ODICAL ROOM
ERAL LIBRARY
NIV. OF MICH.

FEB 18 1924

The Journal of Educational Psychology

Devoted Primarily to the Scientific Study of Problems of Learning and Teaching.

BOARD OF EDITORS:

HAROLD ORDWAY RUGG, *Chairman.*
Lincoln School of Teachers College.
Teachers College, Columbia University.

JAMES CARLETON BELL,
Brooklyn Training School for Teachers.

FRANK NUGENT FREEMAN,
University of Chicago.

ARTHUR IRVING GATES,
Teachers College, Columbia University.

VIVIAN ALLEN CHARLES HENMON,
University of Wisconsin.

RUDOLF PINTNER,
Teachers College, Columbia University.

BEARDSLEY RUML,
Carnegie Corporation, New York City.

LEWIS MADISON TERMAN,
Leland Stanford University.

EDWARD LEE THORNDIKE,
Teachers College, Columbia University.

LAURA ZIRBES, *Assistant Editor.*
Lincoln School, of Teachers College.

VOL. XV.
No. 2.

FEBRUARY, 1924

$4.00 a year.
60c. per copy.

CONTENTS

INTELLIGENCE OF THE FIRST GENERATION OF IMMIGRANT GROUPS. *Gustave A. Feingold* 65

MENTAL DISCIPLINE IN HIGH SCHOOL STUDIES (*Continued*). *E. L. Thorndike* 83

A STUDY OF EXTREME CASES OF UNRELIABILITY. *Percival M. Symonds* 99

PART AND WHOLE METHODS OF LEARNING. *H. B. Reed* 107

A NOTE ON THE SHORTENING OF EXAMINATIONS. *Donald A. Laird* 116

NOTE CONCERNING A T-SCALE FOR WILSON LANGUAGE ERROR TESTS. *G. M. Wilson* 118

NOTES ON ARTICLES IN EDUCATIONAL PSYCHOLOGY IN CURRENT ISSUES OF OTHER MAGAZINES. 120

NEW PUBLICATIONS IN EDUCATIONAL PSYCHOLOGY AND RELATED FIELDS OF EDUCATION. 123

Published Monthly Except June to August by

WARWICK and YORK, Inc.,

York, Pa. *Baltimore, Md.*

ed as Second Class matter November 15, 1921, at the Post Office at York, Pennsylvania, under the Act of March 3, 1879.

The Journal of Educational Psychology

is published monthly, by Warwick & York, Inc., Baltimore, Md., and York, Pa. The issues for 1922 make Volume XIII. Title page and index are bound in December number.

Manuscripts for publication, books or other materials for review, and news items should be addressed to Harold Ordway Rugg, Chairman of the Editorial Board, 59 Edgecliffe Terrace, Yonkers, New York.

The price of the Journal is $4.00 a year, payable in advance. Foreign postage is 30 cents extra. Single issues are 60 cents each, and less than a full year's subscription is charged at single issue rate for every month ordered.

Subscribers should notify the publishers of change in address at least four weeks before publication of the issue with which change is to take effect. No claim for non-receipt of an issue will be entertained unless made within two weeks after receipt of the next succeeding issue.

Warwick & York. Inc.,
Baltimore, Md.

TEACHERS' HEADQUARTERS

European

Plan

Entirely

Fire-Proof

Centrally

Located

EDWARD DAVIS
Manager

HOTEL RENNERT
BALTIMORE, MARYLAND

UNIVERSITY OF VIRGINIA
SUMMER QUARTER

First Term, June 16–July 26
Second Term, July 28–August 30

Courses for Elementary Teachers
Courses for High School Teachers
Courses for College Credit

The Summer Quarter is an integral part of the University year, the courses being the same in character and credit value as in the other quarters of the year.

Degrees are conferred upon men and women for summer work.

The Master's Degree may be obtained by properly qualified students in three Summer Quarters.

It offers opportunities unexcelled in the South and makes a strong appeal to teachers seeking broader scholarship and training and wider social contacts, and to college students desiring to complete degree requirements.

Attendance last Quarter, 2581 from thirty-five states and foreign countries.

The most beautiful and unique educational plant in America.

Accommodations at reasonable rates. Tuition for non-Virginia students $20.00 per term.

Entertainments, Music Festival, excursions.

Reduced rates applied for.

Teachers registered in the Summer Quarter this year will have an excellent opportunity to attend the meeting of the N. E. A. in Washington, to which special excursions will be run.

For illustrated folder and full announcement, write to

SECRETARY, SUMMER QUARTER
UNIVERSITY, VIRGINIA.

State Maintenance of Teachers in Training

BY

WALTER SCOTT HERTZOG

Teaching has failed to attract a sufficient supply of recruits to make adequate selection and training possible. State control of education locates the responsibility for teacher training and for the equalization of educational opportunity upon the state. The policy of training teachers after they enter the service as a substitute for institutional preparation has not been successful, especially in the open country. The unequal salaries paid in the rural schools and in the cities have drawn the better teachers to the centres of population. The public attitude toward teaching has reduced the attendance in teacher training institutions. To improve the attractiveness of teaching as a profession, the state must use all the plans that experience has proved to be helpful. Additional efforts, especially of an economic and social nature, must be made if the rural schools are to be taught by competent teachers. The purpose of this present study is to investigate the principles, problems, and practices involved in a system of subsidies for prospective teachers as one method of recruiting the profession.

144 *pages.* $1.60. *Postage* 8c.

Warwick & York, Inc.,
Baltimore, Md.

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume XV | February, 1924 | Number 2

INTELLIGENCE OF THE FIRST GENERATION OF IMMIGRANT GROUPS

A Study and a Critique

GUSTAVE A. FEINGOLD

Hartford Public High School

Intelligence testing in the United States Army seems to have set in motion two important movements: One scientific and promising much good, the other popular and promising much harm. The one deals with the further development of tests for use in schools; the other concerns itself with racial psychology, and is, willingly or unwillingly, building up a caste among races, by harping on the mental "backwardness" of those nationalities who rated low in the army tests. Needless to say the former movement has been taken up by psychologists and educators in an earnest, scientific manner for further investigation. The latter has been seized upon by feature writers and self-styled social philosophers for the sole purpose of enhancing their own notoriety.

There is a wide gulf between the cautious statements made by the army psychologists concerning the data which they have so ably presented in Vol. XV, "Memoirs National Academy of Sciences," and the unwarranted inferences drawn from these same data by those who have been trying to exploit them for so-called sociological purposes.

Thus in commenting on the noticeable inprovement in intelligence among foreign born groups of different lengths of residence in the United States, the army psychologists very properly say, "It is not possible to state whether the difference is caused by the better adaptation of the more thoroughly Americanized group to the situation of

the examination, or whether some other factor is operative."[1] Other writers, however, take these same data as indicative of a decline in intelligence of the immigrants who have been coming to our shores in the past 25 years.[2] So wholly absorbed in their pet theories are these latter writers that some of them, of even scientific status, like Dr. Carl Brigham, have come to the astonishing point of disregarding the facts which do not support their conclusions, though these facts may have been gathered by various competent investigators in different parts of the country, and at different times.

A good illustration of this appears in "A Study of American Intelligence" wherein the author dismisses the findings of Miss Murdoch, concerning the intelligence of Jewish school children, with the statement that the American children whom they equalled in mentality must have been drawn from an inferior social group. On the other hand, by making the mere assumption that his "army sample of immigrants from Russia is at least one-half Jewish" and that the samples selected as Alpine "is from one-fifth to one-fourth Jewish," he feels himself justified in stating that there is no truth in the popular belief that the Jew is highly intelligent.[3] According to Dr. Brigham's point of view it apparently does not matter that Miss Murdoch's findings in New York City were verified by an investigation conducted in Youngstown, Ohio,[4] wherein it was found that there, too, the Jewish children had exactly the same intelligence quotient as the American white children.

It was for the purpose of finding out whether the mental differences among the racial groups of the first generation are as great as those among the original immigrants of the same races, and also to determine, with as much accuracy as is possible, the actual mental status of the Jewish youth, that the writer undertook the following investigation. The data were gathered at the Hartford Public High School, which at that time, was the only secondary school in the city, and, therefore, enrolled all the high school pupils of Hartford, some 4000 in number. The city itself is fairly cosmopolitan; its population is about 150,000. Its industries are distributed into three main

---

[1] *Op. cit.*, p. 704.

[2] Brigham, Carl C.: "A Study of American Intelligence," pp. 177, 197. Also Young, Kimball: Intelligence Tests of Immigrant Groups. *Scientific Monthly*, Vol. XV, No. 5, pp. 430 ff.

[3] *Op. cit.*, pp. 187 ff.

[4] Pintner, Rudolph, and Keller, Ruth: Intelligence Tests of Foreign Children. *Journal Educational Psychology*, Vol. XIII, No. 4, p. 214.

groups: insurance, mercantile, and manufacturing of a higher type. About 40 per cent of its population is American white, about 20 per cent Jewish, some 12 per cent Irish, 9 per cent Italian, about 5 per cent Scandinavian, 3 per cent Polish, 3 per cent German, 3 per cent English and Scotch, 3 per cent Colored, and the rest is distributed among various other nationalities. Concerning the economic status of the 10 racial groups, treated in this article, it may be said that they are slightly better off financially than the same groups in other cities of the same size.

The mental difference between the native white and foreign born draftees in the Army was found to be about two years. This is also the mental difference between foreigners of over 20 years residence in the United States and those of less than 5 years residence. The question is whether this difference is native or acquired, or whether it is due to a combination of both these factors.

Extremists claim that it is native. The difference in mentality between the longer and lesser resident foreign groups, they say, shows a decline in the mentality of the more recent immigrants.[1] Of course, proper scientific procedure would require that such a conclusion be deferred for at least 15 years; that is to say, until those who were only 5 years or less resident in the United States at the time the psychological tests were given in the army, had been inured to American influences for at least as many years as the group that had been resident 20 years or more, when tested. But the conclusions drawn, besides being scientifically rash, are absurd on their face value for two reasons: First, because no consideration whatever was taken of the immensely varied environmental conditions that prevailed in the different European countries prior to the War; and second, because no cognizance was taken of the fact that, in a general way, the immigrants of 5 years and less residence were hereditarily related to those of 6 to 10 years residence in the United States, and these again were similarly related to those of 11 to 15 years residence, and so forth. In other words, anyone who is at all acquainted with American immigration knows, or ought to know, that the immigration from any one country constitutes a continuous consanguineous stream. To assume, therefore, that the Russians or Italians who immigrated

---

[1] Brigham: *op. cit.* Young, Kimball: *op. cit.* Ward, Prof. Robert De C.: Some Thoughts on Immigration Restriction. *Scientific Monthly*, Vol. XV, Nos. 4 and 6.

to the United States during the 5-year period immediately preceding the outbreak of the World War, were mentally inferior to the Russians or Italians who entered the United States in the '90s, is to assume that a sudden mental deterioration had taken place in European countries at 5-year intervals. Such assumption, of course, would be absurd, and yet it is no more so than the conclusions drawn by writers on immigration who have been using the results of the psychological tests in the Army for the purpose of proving that continued immigration from East European countries will ruin the United States.

It was felt by the writer, therefore, that a more accurate study of the mental differences of races could be made if the unknown factors of foreign birth and foreign upbringing were eliminated, by taking American born samples of various immigrant groups; especially samples that were brought up in the same city and had received the same amount of public education. As we have said, the Hartford High School was an ideal place for the collection of our data, because at the time of this study it enrolled all high school pupils in the city, so that even the factor of social status was neutralized.

Intelligence testing has been conducted at this High School for the past four years. A modified form of the Army Alpha, adapted by the writer especially for high school pupils, is employed. The tests are given to all freshmen upon entering school and, two years later, a second test is given to all the individuals who become lower juniors. In this study, in order to make it as complete as possible, the intelligence scores attained by the seniors in their senior year, are also included.

A percentage scale of intelligence has been developed for each of the four high school classes, and, beginning with 18 Alpha points, corresponding to the mental age of 10 years, a scale of 15 points for each year's difference in mentality has been adopted.[1]

This study enbraces the following racial groups: English and Scotch, considered as one, both because of their smallness of number and also because of their closeness in race and culture; native born Americans; Jews; Germans; Danish and Swedish, considered as one for the same reason as the English and Scotch; French; Irish, those who had at least one parent born in Ireland; Polish; Italian and Colored. These racial groups were studied separately in the freshman, junior, and senior classes. As no mental tests are given in the sopho-

[1] Cobb, Margaret V.: The Limits Set to Educational Achievement by Limited Intelligence. *Journal of Educational Psychology*, Vol. XIII, No. 8, p. 452.

more year, the grades attained by the sophomores when they were freshmen were included with ***bona fide*** freshmen.[1]

Employing Miss Cobb's Conversion Table, the Alpha Scores of the freshmen and juniors were converted into IQ's and the different racial groups treated in each of these two classes were arranged in rank according to their IQ.

Inasmuch as it is held by most psychologists that mental growth ceases at the age of 14½—certainly at the age of 16, which would be just about a half year under the average age of high school juniors—

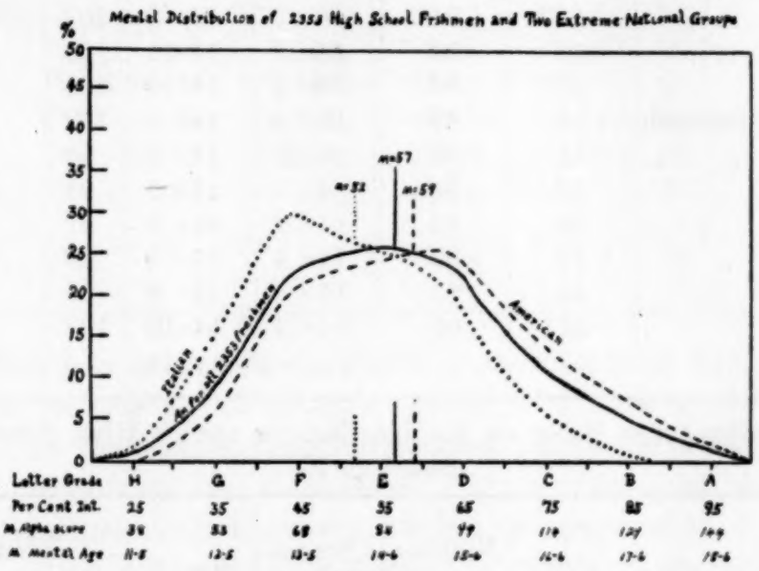

FIG. 1.

the writer deemed it best to arrange the racial groups in the senior class according to the absolute number of points scored, irrespective of chronological age.

The results found among the freshman, junior and senior classes are presented in Tables I, II, and III. Due allowance should be made, of course, for difference in numbers. We cannot say that the 35 children of French extraction are as representative of French mentality as the 892 American children are representative of American mentality.

The extent of overlapping between the American and Italian freshman groups is presented graphically in Fig. 1.

---

[1] The freshman intelligence scores were used of the classes which entered January, 1923, September, 1922, January, 1922 and September, 1921. The junior scores were used of the classes which entered January, 1921 and September, 1920. The senior scores were used of the classes which entered January, 1920, September, 1919 and January, 1919.

TABLE I.—RELATIVE RANK OF RACIAL GROUPS AMONG HIGH SCHOOL FRESHMEN

Based on IQ

Nationality	Per cent intelligence	Number of points	Chronological age	Mental age	IQ	Number of cases	Per cent of total
English and Scotch..	60	95	14- 5	15- 2	105	76	3.2
American (native born).....	59	94	14- 7	15- 0	103	892	37.9
Jewish.....	58	93	14- 5	14-11	103	518	22.0
German.....	57	91	14- 5	14-10	103	86	3.7
Danish and Swedish	56	89	14- 6	14- 9	102	114	4.8
French.....	55	86	14-10	14- 7	98	35	1.5
Irish.....	53	84	14- 7	14- 5	98	278	11.7
Polish.....	52	82	14- 9	14- 3	97	90	3.7
Italian.....	52	82	14- 9	14- 3	97	206	8.7
Colored.....	54	85	15- 4	14- 6	95	58	2.5
Average.....	57	91	14- 7	14-10	102		
Total.....	..	..	.....	.....	...	2353	99.7

TABLE II.—RELATIVE RANK OF RACIAL GROUPS AMONG HIGH SCHOOL JUNIORS

Based on IQ

Nationality	Per cent intelligence	Number of points	Chronological age	Mental age	IQ	Number of cases	Per cent of total
English and Scotch..	69	128	16-0	17- 5	109	17	2.6
American.....	68	126	16-5	17- 4	107	286	43.0
German.....	66	122	16-2	17- 0	105	20	3.0
Jewish.....	65	121	16-4	16-11	103	208	31.3
Danish and Swedish.	62	115	16-4	16- 6	101	28	4.2
French.....	..	...	....	.....	...	6	0.9
Polish[1].....	62	115	16-2	16- 5	101	11	1.7
Irish.....	61	113	16-5	16- 5	100	44	6.6
Italian.....	62	115	16-5	16- 6	100	29	4.4
Colored.....	57	105	17-8	15-10	94	15	2.3
Average.....	66	122	16-5	17- 0	103		
Total.....	..	...	....	.....	...	664	100.0

[1] The large shrinkage in the percentage of Polish children enrolled in the junior year indicates that their rise in mental rank is due to positive selection—the inferior individuals having dropped out in larger proportions than among the other racial groups. This is borne out by the fact that the average chronological age of the Polish group in the above table is comparatively low.

TABLE III.—RELATIVE RANK OF RACIAL GROUPS AMONG HIGH SCHOOL SENIORS Based on Alpha Points Scored

Nationality	Per cent intelligence	Number of points	Chronological age	Mental age	IQ	Number of cases	Per cent of total
English and Scotch..	74	142	17–10	18– 4	103	19	3.3
German.....	72	139	18– 0	17–10	99	13	2.3
American.....	70	136	17–11	17–11	100	264	46.5
Danish and Swedish.	67	131	17–10	17– 6	98	30	5.3
Jewish.....	64	125	17–10	17– 1	96	146	25.7
Irish.....	64	125	17– 7	17– 1	97	53	9.3
Colored.....	63	121	18– 7	16–10	91	9	1.6
Italian.....	61	118	17– 4	16– 8	96	16	2.8
Polish.....	60	117	18– 0	16– 7	92	11	1.9
French.....	..	...	.....	.....	...	6	1.1
Average.....	67	130	17–10	17– 6	98		
Total						567	99.8

Confining our attention to Table I we note that the differences among the various racial groups, either in points scored or in IQ, diminish from the highest (English and Scotch) to the lowest, Colored, by almost imperceptible differences. It is true that the Colored individuals, considered in these tables, were favored by positive selection. That is to say, only those Colored boys and girls who, in grammar school, gave promise of success were sent on to high school. Unquestionably a larger proportion of poor material among the Colored boys and girls leave grammar school to go to work than among the other racial groups. Instead of having racial differences, therefore, we may say that there is a slight difference between groups of races, and, from Table I, we would be justified in making three divisions; placing in the first group the English, American, Jewish, and German; in the second, Scandinavian, French, and Irish; and in the third, the Polish and Italian. The Colored individuals, of course, would have to constitute a group by themselves. But even here it is difficult to say whether the difference is native or whether it is due to a difference of cultural tradition and educational opportunities. Notwithstanding that the 2353 individuals treated in Table I are all grammar school graduates and most of them have been brought up in the same city, we nevertheless cannot disregard the fact that most of the Jewish and nearly all of the Polish and Italian children were

somewhat handicapped in the modified Alpha Tests because English is not the spoken language in their homes.

A second fact to be noted in Table I is that the mental difference between the American born and the lowest ranking white foreign race is only nine months. Compare this with the mental difference found between the native and foreign born United States draftees, which was about 2 years. If the first generation of East European immigrants can more than half bridge the mental gap that exists between their parents and native Americans, is it not more than probable that most of the original gap was due to environment, *i.e.*, differences in educational opportunity, rather than to heredity?

But this fact is significant; namely, that when the groups become more select as they reach the lower junior year and have, therefore, become still more initiated to the same environmental and educational influences, the relative difference among them is still the same as in the freshman year; although it should be noted that it is an entirely different group of individuals that was tested. Only two racial groups change position in the junior year—the Jewish and German—and even here the difference in points scored, or IQ, is comparatively small.

Though it is true that with the one exception of the Colored pupils, selection plays little or no rôle in the freshman class, still it cannot be said that economic selection does not begin to function in the junior year. This is proved by the fact that although the Irish constitute 11.7 per cent of the entire enrollment in the freshman year, they constitute only 6.6 per cent of the enrollment in the junior year. Similarly the proportional withdrawals during the first two years of high school attendance among the Italian and Polish children are almost twice as large as those of American, or Jewish, or Danish children. Hence, in the junior year we should expect the children of Irish, Polish, and Italian extraction to make a better showing than in the freshman year. To be sure the eleven individuals of Polish birth *do* climb up one degree, but much significance cannot be attached to this because of their fewness.

In the senior year, both scholastic and economic selection having played a still greater rôle, we notice a shifting in rank among several of the racial groups. Unfortunately, however, some of these racial groups such as the German, Colored, Italian, and Polish, have so few representatives in this year, that these results are considerably less reliable than those for the freshman and junior years.

The writer deemed it advisable, however, to represent the results for the racial groups in the three classes separately rather than to combine individuals of the same race from the three classes. His hope was that they would line up exactly the same way each time. And, as a matter of fact, they do so in the freshman and junior years with but one slight change. Also they show the same tendency in the senior year.

The best way to obtain a true estimation of the racial differences, however, is to combine the ranks attained by the racial groups in each of the three high school classes and arrange them according to the average rank.[1] This is done in Table IV.

TABLE IV.—RELATIVE RANK OF RACIAL GROUPS AMONG HIGH SCHOOL PUPILS BASED ON IQ FOR FRESHMEN AND JUNIORS AND POINTS SCORED FOR SENIORS[1]

Nationality	Freshman rank	Junior rank	Senior rank	Sum of ranks	Average rank	Average deviation
1. English and Scotch	1	1	1	3	1.0	0.00
2. American (native born)	2	2	3	7	2.3	0.43
3. German	2	3	2	7	2.3	0.43
4. Jewish	2	4	5	11	3.7	1.10
5. Danish and Swedish	5	5	4	14	4.7	0.43
6. French	6	.....	.....	.....	.....	.....
7. Irish	6	7	5	18	6.0	0.67
8. Polish	8	5	9	22	7.3	1.57
9. Italian	8	7	8	23	7.7	0.43
10. Colored	10	9	7	26	8.7	1.10

[1] The table reads: In the freshman year the English and Scotch ranked first; the native born Americans, the Germans and Jews tied for second place; the Danish and Swedish ranked fifth, etc.

The combined rank presented in Table IV is so close to the individual ranks shown in Tables I and II, and these again, are so fairly harmonious, though the results were obtained with entirely different individuals, that there can be but little doubt that the relative rank

[1] The writer is aware that the main objection to this method is that the same weight is given to averages attained by groups that differ considerably in size.

of the different racial groups shown in Table IV represents the true situation concerning their mental alertness as measured by verbal intelligence tests.

Now, how do these ranks compare with the ranks occupied by the same races in the Army Tests? To be sure the French and Jewish do not appear as distinct racial groups in the Army results. But if we select each of the races herein considered from those presented in the Army tables, including the Russians because they may contain some Jews and some Poles, and leaving each of these races exactly in the order in which they appear with respect to one another, we obtain Table V, which gives their Army Ranks based on four different points of reference, and finally on the average combination of these four.

TABLE V.—RELATIVE RANK OF RACIAL GROUPS AMONG U. S. DRAFTEES

	Rank based on per cent scoring A and B	Rank based on per cent scoring D, D−, and E	Rank based on per cent at or above average of officers	Rank based on position on combined scale	Sum of ranks	Average rank	Average deviation
1. English	1	1	1	1	4	1.0	0.00
2. Scotch	3	2	2	2	9	3.0	0.75
3. American (native born)	2	5	3	4	14	3.5	1.00
4. German	4	3	4	3	14	3.5	0.50
5. Danish	5	4	7	5	21	5.2	0.85
6. Swedish	6	6	5	6	23	5.8	0.35
7. Irish	7	7	6	7	27	6.8	0.35
8. Russian	8	8	8	8	32	8.0	0.00
9. Italian	9	9	9	9	36	9.0	0.00
10. Polish	11	10	10	10	41	10.2	0.35
11. Colored	10	11	11	11	43	10.8	0.27

It is astonishing to note the uniformity of results obtained by the Army psychologists with United States draftees and by the writer with high school pupils. Whether this uniform difference in rank among the ten racial groups is indicative of native endowment or of the same combination of environmental influences operating among the American born descendents in the same manner as among their foreign born parents, is difficult to say. The final rank of the racial groups as obtained by the Army psychologists with draftees and by the writer with high school pupils is given in Table VI.

TABLE VI.—FINAL RANK OF RACIAL GROUPS AMONG U. S. DRAFTEES (AVERAGE AGE 25–0), HIGH SCHOOL FRESHMEN (AVERAGE AGE 14–6) AND HIGH SCHOOL FRESHMEN JUNIORS AND SENIORS COMBINED

ARMY	FRESHMEN	THREE CLASSES COMBINED
1. English	1. English and Scotch	1. English and Scotch
2. Scotch	2. American	2. American
3. American	3. Jewish	3. German
4. German	4. German	4. Jewish
5. Danish	5. Danish and Swedish	5. Danish and Swedish
6. Swedish	6. French	6 ? (French)
7. Irish	7. Irish	7. Irish
8. Italian	8. Italian	8. Polish
9. Polish	9. Polish	9. Italian
10. Colored	10. Colored	10. Colored

The important conclusions thus far derived from this study may be summed up as follows:

1. The mental differences among the American reared descendants of foreign races, viewed among themselves, are so small as to be practically insignificant.

2. The mental difference between the American reared children of foreign races and children of Anglo-Saxons, whether judged by the freshman or by the junior scores, is equally too small to be significant. These two facts prove, all jingoism to the contrary, that if their influence is exerted throughout the period of native mental growth American institutions, or at least the American public schools, do constitute a very decided melting pot.

3. The Jewish children of high school age, though subject to less positive selection than pupils of the other racial groups, have practically the same mental rank as the American children in their freshman year, and are only a few points lower than they in their junior year. Inasmuch as this finding harmonizes with the results obtained by Miss Murdoch and Dr. Pintner, there seems to be more truth in the popular opinion concerning the high intelligence of the Jews than there is in Dr. Brigham's scientific assumption concerning their lack of intelligence. Hence, the cry from the heart raised by Professor Albert Bushnell Hart that "Where they [the immigrants, meaning Jews] learn the English language, their children press hard against those of Colonial stock and later immigrant races, in schools, and colleges and trades,"[1] contains considerable truth. To verify it one need but con-

[1] Hart, Albert Bushnell: New Race of the Races. *New York Times*, Sept. 16, 1923.

sult the honor roll of any cosmopolitan school or college. But whether this cry is manly or babyish is another question.

4. In view of the fact that the difference in points scored between the high ranking and low ranking racial groups increases after the sixteenth year, it would seem that this is due to increased differences in social opportunities met by the children of the different races after they have reached the adolescent stage. Thus the absolute difference between the average Alpha points scored by the English and Scotch on the one hand, and the Italians on the other, for the freshman and junior years respectively is only 13 points, whereas in the senior year it becomes 24 points. Unquestionably if these same individuals were to be given an intelligence test 10 years after graduating from high school, still greater differences would be found among the various racial groups. The reason is that the English, American, German, and Scandinavian children would doubtless be able to continue to progress materially, socially, and professionally, without any undue hindrance on account of race or creed; whereas children bearing Jewish, Polish, Italian, and, in some localities, Irish names would find some opportunities for development closed to them. The only alternative to the above explanation is that different races differ with respect to the age when native mental growth ceases, and that Anglo-Saxons continue to develop beyond the age of 14 or 16. No evidence of this, however, has thus far been found. Though we have proved up to this point that the mental difference between the first generation of foreigners and native born Anglo-Saxons is so small as to make the flinging of psychological stones at immigrant groups seem ridiculous, this study would not be complete if it were to stop with an essay of groups merely from the point of view of mental alertness. It seems that practically every popular writer and no few scientists, who have been using the results of the Army Tests as a weapon against foreign races, have been laboring under the impression, or at least have given their auditors the impression, that there is nothing else in the psycho-physical make-up of the individual except mental alertness. Yet the very corner stone of the reliability of intelligence tests rests upon their correlation with achievement which involves the exercise of general intelligence. The Army Tests were validated, in part, by the fact that officers of the regular Army did excellently well in them, while mentally backward individuals (having been determined to be such by other devices) did astonishingly poorly. Also all intelligence

tests correlated if not high, at least positively, with grammar, high school and college grades; and certainly intelligence is needed for successful achievement in studies.

However, to explain the phenomenon of individuals scoring high in the verbal intelligence tests who do miserably poor work in scholarship, the psychologists defend their position by saying that these individuals do not exert a normal amount of effort, application, or industry in their studies. On the other hand, to explain the phenomenon of individuals who score low in intelligence tests yet do excellent work in school or college, the statement is made that either there was an unknown slip-up in their cases, at the time of the tests, or else they are putting forth abnormal efforts in their studies. In either case it is admitted that in addition to mere mental alertness, another element in the human make-up is necessary for success, namely, effort, application, exertion of the will. Now, before belaboring immigrant groups for their apparent lack of mental ability, as measured by verbal tests, it would seem to be required by the scientific temperament to determine first whether what these immigrants lack in mental alertness is not more than made up for by tenacity of will and constancy of purpose.

To be sure, there are, as yet, no definite and reliable means of measuring this element among racial groups, on a large scale. The only thing to do is to make general observations and to draw such conclusions from them as seem warrantable. Thus if it is found by manufacturers that certain racial groups show more faithfulness and greater tendency to stay at their employment than others, assuming that it is equally lucrative to all, it is fairly safe to conclude that they constitute racial groups which are endowed with a stronger will-power than the others. Again, if it is observed by school officials and investigators that the children of certain racial groups tend to persist in school longer than others, economic and mental conditions being the same, then it is safe to assume that these individuals as a race possess a stronger tenacity or will purpose.[1]

Since will-power or effort is a very essential element for success in studies, we can determine how different races rank in this element by comparing various racial groups of exactly the same mental level

---

[1] J. K. Van Denburg found, for instance, that children of Jewish descent tend to persist in high school in larger proportion than other racial groups ("Causes of the elimination of students in Public Secondary Schools in New York City").

and thus cancelling out the factor of mental alertness. The objective measurement of will-purpose, or effort, will then be obtained by the formula, Effort = $\frac{\text{Scholarship}}{\text{Intelligence}}$. Thus, if a pupil who possesses 60 per cent intelligence obtains 60 in Latin, we infer that he is putting forth a normal amount of effort in this study. If, on the other hand, the same pupil obtains a scholarship grade of 40 in Algebra, then we find that he is putting forth an amount of effort measured by the fraction 40/60 which equals 67 per cent. Finally, should his scholarship in some subject be 85, then the degree of effort put forth in that subject would be 125, or considerably above normal. And this is the way effort should be measured, for while mental alertness or intelligence may be limited, there is no limit to the amount of effort a person may put forth in any undertaking. The effort or will-power shown by different individuals or racial groups can be measured in the same manner of course, other things being equal.

The writer has attempted to achieve this task with high school freshmen of American, Irish, and Jewish descent. Final examination marks were used as the measurement of scholarship in order to insure absolute impartiality and to eliminate the possibility of a racially influenced subjective judgment on the part of the teacher. Other racial groups were not included because there were not enough cases. Both the intelligence and scholarship marks were reduced to terms of percent. Various mental levels of the three racial groups were then made up by including in each mental level all those individuals of a specific race who possessed an intelligence mark corresponding to that level. No mental level of any racial group was made up unless there were at least three individuals in it. To assure further accuracy, differences in effort among the racial groups that might be due to a preponderance of individuals of one sex or the other, were eliminated by dividing the 588 freshmen involved in this study into boys and girls, on one hand, and by further dividing them into academic and non-academic groups. The effort of the American, Irish, and Jewish children was studied in first-year high school Algebra and English because these two subjects are compulsory for all freshmen, and also because all pupils, whether academic or non-academic, have the same teachers, use the same texts, and have exactly the same assignments in them. Table VII shows the effort quotient, (EQ) obtained by the three racial groups in beginners' high school algebra.

TABLE VII.—RACIAL EFFICIENCY QUOTIENTS IN ALGEBRA[1]

Mental level.....	A	B	C	D	E	F	G	Average of individual EQ's or REQ	Per cent of pupils in racial groups
Academic Girls (85 cases)									
American.....	..	..	82	87	85	92	....	84	54.0
Irish.....	..	..	56	....	95	....	....	(73)	10.5
Jewish.....	..	..	97	86	110	113	....	95	35.5
All others.....	..	..	..	....	....	....	....	...	0.0
Non-academic Girls (225 cases)									
American.....	..	..	71	88	102	83	124	99	37.5
Irish.....	..	..	60	69	67	....	113	80	13.6
Jewish.....	..	..	84	83	94	85	134	93	27.2
All others.....	..	..	..	....	....	....	....	...	22.0
Academic Boys (112 cases)									
American.....	..	74	66	83	82	41	....	74	46.0
Irish.....	..	..	73	86	....	....	....	(78)	8.0
Jewish.....	..	75	81	90	96	65	....	83	34.5
All others.....	..	..	..	....	....	....	....	...	11.5
Non-academic Boys (166 cases)									
American.....	..	67	71	50	77	76	95	71	50.0
Irish.....	..	..	34	75	101	68	123	70	13.5
Jewish.....	..	56	94	101	85	128	....	100	9.3
All others.....	..	..	..	....	....	....	....	...	27.0

[1] Regarded objectively EQ is the efficiency quotient; subjectively, it is the effort quotient.

Here we find that out of 19 mental levels the Jewish children rank first in 14 of them in the amount of effort put forth in the study of Algebra. It should be constantly borne in mind, in viewing these figures, that the element of intelligence, or mental alertness, is entirely

neutralized, inasmuch as we are comparing racial groups of exactly the same mental level in each case.

There are two possible objections that may be raised by the person who, in spite of the evidence thus far presented, is still convinced that the East European races are mentally inferior to the so-called Nordic races. He may say that the Jewish children in the C or D or E or any mental level, were under-rated in the tests, and that many of them doubtless belonged to a higher level, in which case, their effort, or efficiency quotient, would have been smaller than it actually is. But if he makes such a statement he gives away the entire case of the die-hard, genetic mental testers. For it would be a confession that verbal intelligence tests under-rate the mental ability of individuals of foreign extraction. The other objection that he may raise is that our figures indicate, what has been generally observed, that Jewish children are especially gifted in mathematics. It is necessary, therefore, to go to our next table which shows the effort quotients of the same three racial groups in beginners' high school English.

Here, too, we find that the Jewish children rank highest in effort in 11 out of 19 cases, and tie with the Irish children in one case. Certainly their superior efficiency in English cannot be attributed to special ability as it may be in the case of mathematics.

These figures are not presented for the sake of showing either the virtues or defects of any particular race. They would prove the same thing if either the Irish or American group made the same showing. They are given here merely for the purpose of demonstrating, as objectively as is possible, that mental alertness is not the entire be-all and end-all of the individual and the race. Whether or not, intelligence being the same, the superior efficiency in Algebra and English, of the racial group above mentioned is due to greater interest in book learning than manifested by the two other racial groups, does not affect the issue. The point the writer wishes to make is that there are other psychophysical elements which function as indispensable co-partners of mental alertness in the accomplishment of those things which are attributed to general intelligence.

Of course, the man on the street will immediately say that all the children of Jewish birth are "grinds" and everybody knows that anyway. But if "grinding" at the useful task in which the individual is engaged is not a manifestation of general intelligence, what is? As a matter of fact, that which measures the intelligence of a nation is its ability to pursue and to consummate remote and far-reaching

TABLE VIII.—RACIAL EFFICIENCY QUOTIENTS IN ENGLISH

Mental level.....	A	B	C	D	E	F	G	Average of individual EQ's or REQ	Per cent of pupils in racial groups
Academic Girls (85 cases)									
American.....	..	..	95	100	100	104	....	98	54.0
Irish.....	..	..	79	....	105	....	....	(92)	10.5
Jewish.....	..	..	105	105	127	119	....	106	35.5
All others.....	..	..	....	....	....	....	....	...	0.0
Non-academic Girls (225 cases)									
American.....	..	..	94	98	116	110	142	108	37.5
Irish.....	..	..	95	90	113	....	140	105	13.6
Jewish.....	..	..	104	95	107	130	141	114	27.2
All others.....	..	..	....	....	....	....	....	...	22.0
Academic Boys (112 cases)									
American.....	..	79	77	82	94	91	..	84	46.0
Irish.....	..	..	96	94	....	....	....	(95)	8.0
Jewish.....	..	83	81	94	108	94	....	90	34.5
All others.....	..	..	....	....	....	....	....	...	11.5
Non-academic Boys (166 cases)									
American.....	..	78	81	84	82	89	110	84	50.0
Irish.....	..	..	62	93	82	109	109	87	13.5
Jewish.....	..	65	83	83	80	128	....	90	9.5
All others.....	..	..	....	....	....	....	....	...	27.0

plans which affect the national life. The state undertaking a national policy or a piece of work which requires 10 or 100 years to complete, gives proof, not only of ability, but of far-sightedness, of strong will-power, of unity of purpose and, in fact, of every attribute which makes for national conservation; and, be it remembered, conservation or survival in terms of the biologist, is the ultimate proof of general

intelligence. Now, what is true of the race is equally true of the individual, barring the pathological case, of course. Proficiency in letters, in science, in arts, in invention, in industry, and in commerce all require not only mental alertness and such elements as memory, ingenuity, reasoning ability, and the like, but also strength of will, sheer tenacity and dogged stick-to-itiveness.

It is true that these qualities without mental ability back of them cannot lead to anything, but it is equally true that mental alertness without these qualities also can lead nowhere. In the future psychological surveys of immigrant groups, therefore, it is to be hoped, for the sake of clear public thinking and for the safe-guarding of our internal American amity, that investigators and writers will take more than a half-view of the psychophysical make-up of alien races.

MENTAL DISCIPLINE IN HIGH SCHOOL STUDIES

(Continued from January)

E. L. THORNDIKE

With the aid of the staff of the Institute of Educational Research

Teachers College, Columbia University

The effect of study itself is not, in Tables IX and X, isolated from its effect through its associates. If French and chemistry always were taken together, the superiority in gain shown in Table X for both will relate to the combined effect of both; and if that of one of the two were zero, it might really belong to one alone. To isolate the effect of a study we need to compare those who took it with those, alike in other respects, who took something other than it. We can get large enough populations to do this by treating some studies as equivalents on the basis of the facts of Table X. For example, History Music, Shop, Spanish, English, drawing, and business are alike in Table X in having very closely an indifferent relation to gain. Even if some one of them, like Business, is there lower than it should be by its association with stenography, and some other like Shop, is higher than it should be by its association with maleness (we shall later see that the boys gained somewhat more than the girls) still there will at the most be no very large differences between one or another study of this group. So we may treat any one of them as Study I and if we find 500 pupils who take Latin, Geometry, French, and Study I, and 500 others who take Latin, Geometry, French, and Chemistry, we may compare the effect of Chemistry with that of Study I.

Following a similar argument, we may treat a course in Civics, Economics, Psychology, or Sociology as a course in Study II. These studies are about alike in Table X, and there is no reason to think that they are greatly unlike in their real intrinsic effects on the ability tested by the examination. We have combined agriculture with Biology as Study III, though the cases of agriculture are so few as to amount to very little.

Arithmetic and bookkeeping are both put as Study IV.

Geometry and Algebra and Trigonometry are taken as Study V.

Latin and French are taken provisionaly as Study VI. These last two amalgamations are more questionable, since Trigonometry is definitely above Geometry or Algebra in Table X, and French is considerably above Latin. Also French and Latin would be considered

at least by many classicists, as very different in the mental training which they give. However, the amalgamation has really only the effect of giving Study V a little better record than Geometry alone would have, and of giving Study VI a little better record than Latin alone would show. Stenography, cooking, sewing, and the combined course in cooking and sewing are the members of Study VIII, similar in representing courses in concrete learning combined with manual dexterity, taken largely by girls, and all at the bottom of Table X. Chemistry, physics, and general science are listed as Study IX. The other subjects are left as they were. We now make all possible comparisons where Study II is in place of Study I, or of nothing. For example, we have 83 pupils taking all I's, and 74 taking one II and the rest I's, 313 pupils taking V, VI, and one or more I's, and 33 taking one II and V, VI, and one or more I's, 214 pupils taking V, IX, and one or more I's and 81 taking one II and V, IX, and one or more I's, and so on. We find the average gain[1] for each group and the weighted average superiority or inferiority in gain of Study II in comparison

TABLE XI.—THE DIFFERENCES IN GAIN BETWEEN PUPILS TAKING II AND PUPILS WHO TAKE I OR NOTHING IN PLACE OF II

An Entry of I under Program May Mean One or Two or Three I's. D = dramatic art. T = physical training. M = manual training.

Program	n_1	Average gain (g_1)	Program	n_2	Average gain (g_2)	$g_1 - g_2$	Weight
II, I	74	26.82	I	83	23.10	+ 3.72	74
II, I, VIII, VIII	52	18.08	I, VIII, VIII	126	16.55	+ 1.53	52
II, I, V, VI	33	35.68	I, V, VI	313	28.91	+ 6.77	33
II, I, V, IX	81	28.86	I, V, IX	214	25.49	+ 3.37	81
II, I, VIII	70	19.71	I, VIII	155	19.47	+ 0.24	70
II, I, V	69	19.75	I, V	264	26.16	− 6.41	69
II, I, IV	32	30.31	I, IV	64	22.73	+10.71	32
II, I, VIII, VIII, VIII	25	15.30	I, VIII, VIII, VIII	67	15.19	+ 0.11	25
II, I, D	9	23.61	I, D	20	24.25	− 0.64	9
II, I, D, T	4	43.00	I, D, T	7	38.14	+ 4.86	4
II, I, M	2	12.50	I, M	6	34.17	−21.67	2

and so on for 134 such comparisons in all, with 1191 as the sum of the weights and +592.92 as the sum of the weighted differences ($g_1 - g_2$), giving +0.5 as the weighted average difference for Subject II in comparison with I or nothing. The probable error of this +0.5 is relatively "large." It may be estimated as 0.6.

[1] Actually the average gain +2½, since for convenience, the gains, Form A to Form B, were taken at their face value and the gains Form B to Form A were increased by 5 to balance. The *superiorities* or *inferiorities* are uninfluenced by this treatment.

with Study I or nothing, each difference being weighted by the smaller of the two populations involved, as shown in Table 11.

Tables similar to Table XI were made, but for Studies III, IV, V, VI, VIII, IX, Physical Training, Dramatic Art, and Manual Training. The grand total results for all these are shown as Table XII. In these computations, subject II was used wherever it could be used, including comparisons like II, II, I, III, with II, I, III. So for all other subjects. The effect of the subject when added to a program already containing it, is less than when it occurs alone, probably in part because such double entries mean only that the pupil reported twice a subject taken in two semesters. An error is thus introduced if such doublets and triplets occur oftener in the case of one subject than another. The error is, however, small; and it would probably be a much worse error to omit all second and third appearances from the computation. For example, cooking when taken with sewing, or with stenography, would then have no showing in our results.

The probable error of the weighted average of Study II to Study I or nothing will be somewhat less than the probable error of a difference due to two averaged gains each computed from a population of 1191 with a variability (mean square deviation) of 22.9. The mean square deviation of the entire group in gain is 22.9. That of a group of pupils all taking similar programs will be probably a little less than 22.9; and, in our pairs of groups, the average population of the two is much higher than the population used in computing the weighted average difference. As a rough approximation, I have taken, as the estimated probable error, .9 (PE of difference with n_1 and n_2 as 1191 each, and SD_1 and SD_2 as 22.9 each). The chief incidence of the error is with the V's and with the VIII's. In the former case the use of only single V's gives +3.17 instead of 3.10 in Table XII; in the latter the use of only single VIII's gives −2.21 in place of −1.51.

If the pupils taking II or III or IV, etc., were in each case a random sampling from all the pupils in the investigation, Table XII would provide measures of the comparative effect of a year of one study rather than another, including whatever was implied by taking one study rather than another.

They are not, however, random samplings. For example, there is reason to believe that, within the same sex, abler pupils take trigonometry and advanced algebra than take stenography, manual training, sewing, and cooking; and common knowledge informs us that more boys than girls take physics and chemistry and more girls

than boys take cooking and sewing. The facts of Table X and of Table XII both show certain features, such as the superiority of shop to cooking and sewing, and of physical training (a "male" subject in these schools), to dramatic art (a "female" subject in these schools); which suggest that maleness is correlated with gains. The fact that the gains in a colored school in City 1 average only about 14.0 while the gains in the white schools of that city average 23.5, and that this inferiority in gain corresponds to a difference between about 120 and 190 in average 1922 score suggests that the abler pupils would gain more even if they took the same programs as the less able. The results of Table XII then may need correction for sex and for initial ability. The facts show that they do.

TABLE XII.—THE DIFFERENCE IN GAINS BETWEEN PUPILS TAKING ANY GIVEN SUBJECT AND THOSE WHO TAKE I OR NOTHING IN PLACE OF IT

SUBJECT	WEIGHTED AVERAGE DIFFERENCE	SUM OF WEIGHTS
II	+0.50	1191
III	−0.85	1375
IV	+2.73	1026
V	+3.10	1720
VI	+2.21	2045
VIII	−1.51	2153
IX	+3.48	1261
D	−1.23	466
T	+0.24	1206
M	−0.35	95

The essential facts for sex are shown in Table XIII. Being a boy rather than a girl taking the same group of studies seems to imply an average superiority in gain of 3.1 points. Whether or not this holds true for boys and girls in high school, in general, it certainly requires that the facts of Table XII be corrected for the probable effect of sex in our groups. This can be done by subtracting 1.55 from all male gains and adding 1.55 to all female gains in all of the comparisons in Table XI and the other similar tables whence Table XII was derived. A rough, but impartial, approximation can be achieved by doing so only for the main comparisons (those with the smaller $n > 19$) and assuming that the minor comparisons will in the long run give similar results. This is what we have done.

The influence of initial score upon gain, for pupils of the same sex taking the same program of studies, may be determined in two ways. First, we may measure the actual obtained relation of initial score to

TABLE XIII.—GAINS OF BOYS AND GIRLS TAKING THE SAME STUDY GROUP

Study group	*n*	Boys average gain	*n*	Girls average gain	Superiority of boys
All I	48	24.82	36	20.88	+ 3.94
I, II, V, IX	71	28.04	10	31.50	− 3.46
I, II, IX	66	26.90	17	27.35	− 0.45
I, III	56	24.94	55	21.82	+ 3.12
I, III, V	74	22.83	29	23.12	− 0.29
I, III, V, VI	55	34.43	66	31.45	+ 2.98
I, III, VIII	10	18.00	76	18.39	− 0.39
I, IV, VIII	18	20.06	84	26.01	− 5.95
I, IV, VIII, VIII	21	23.88	95	22.52	+ 1.36
I, V	214	29.17	50	18.84	+10.33
I, V, T	84	29.18	16	21.88	+ 7.36
I, V, VI	186	28.12	129	28.45	− 0.33
IV, VI, T	81	28.10	52	21.17	+ 6.98
IV, VI, VIII	8	23.38	76	27.80	− 3.42
I, V, VI, IX	82	34.62	47	28.63	+ 6.02
I, V, IX	185	26.40	31	19.14	+ 7.26
I, VI, IX	41	34.60	41	33.16	+ 1.44
I, VIII	19	24.95	83	20.75	+ 4.20
I, VIII, VIII	12	15.58	117	16.14	− 0.56
I, IX	92	31.67	36	26.40	+ 5.27
Weighted average superiority (by smaller *n*).....					+ 2.94
Combination of groups.....	215	29.50	370	25.96	+ 3.54

gain in groups of the same sex and studies, correcting it for the effect produced by imperfections in the measurements. Second, we may compare the average gains of two large groups differing so widely in initial score that the effect of the imperfections in the measurements is trifling. In our case the colored and white pupils will serve this purpose admirably.

The facts for the effect of initial score, the studies taken being constant, and sex also, are shown in Table XIV, column A. These are, however, as stated above, subject to a special effect of the imperfection of the measurements, and will mislead if it is not allowed for. This effect, which has been treated in detail elsewhere, is the distortion of the relation. The median obtained correlation between 1922 score and gain for the 30 groups taken, is $-.20\frac{1}{2}$ ($-.19$ or $-.21$ accord-

TABLE XIV.—THE CORRELATIONS BETWEEN THE 1922 SCORE AND GAIN, AND BETWEEN THE SUM OF THE 1922 AND 1923 SCORE AND GAIN, IN THE CASE OF PUPILS OF THE SAME SEX TAKING THE SAME PROGRAM OF STUDIES (*i.e.* OF THE COMBINED STUDY-GROUPS, I, II, III, IV, V, ETC., OF TABLE XII)

Group	*n*	A *r*('22) (Gain)	B *r*(sum) (Gain)	B − A
1	30	−.59	−.50	.09
2	39	−.56	−.61	−.05
3	76	−.44	−.03	.41
4	56	−.40	+.22	.62
5	71	−.37	−.22	.15
6	37	−.35	.00	.35
7	26	−.35	.00	.35
8	64	−.31	+.15	.46
9	63	−.31	+.08	.39
10	59	−.31	−.02	.29
11	65	−.26	+.12	.38
12	38	−.24	+.10	.34
13	150	−.22	+.15	.37
14	43	−.22	+.11	.33
15	35	−.22	−.40	−.18
16	56	−.22	.00	.22
17	39	−.20	.00	.20
18	99	−.18	−.03	.15
19	40	−.16	+.08	.24
20	42	−.15	+.18	.33
21	84	−.13	+.15	.28
22	148	−.11	+.09	.20
23	48	−.09	+.06	.15
24	65	−.07	+.07	.14
25	76	−.05	+.11	.16
26	39	−.04	+.20	.24
27	62	.00	+.10	.10
28	99	+.08	+.37	.29
29	112	+.14	+.36	.22
30	39	+.12	+.45	.33
Total.....	1863			
Median (equal weights).....		−.22	+.08½	+.26
Median (weights by *n*'s).....		−.19	+.10	+.24
Weighted average of medians....		.....	.....	+.26

TABLE XV.—THE PROCEDURE IN CORRECTING THE FACTS OF TABLE XII FOR INEQUALITIES OF SEX AND INITIAL ABILITY

Study A Group A	Study B Group B	n_A	n_B	Gain for Study II over I or Nothing Due to: Allowance for sex	Allowance for initial score	Combined allowance	(Combined allowance) x (smaller n)
II, I	VII, I	74	83	−.45	− .03	− .48	−35.52
II, I, I	I, I	22	27	−.77	− .55	−1.32	−29.04
II, I, III	I, III	37	112	−.56	− .51	−1.11	−41.07
II, I, III, VIII	I, III, VIII	36	86	−.24	+ .55	+ .31	+11.16
II, I, IV	I, IV	32	64	−.17	+1.65	+1.48	+47.36
II, I, IV, VIII, VIII	I, IV, VIII, VIII	22	120	+ .04	+ .03	+ .07	+ 1.54
II, I, V	I, V	69	264	+ .19	− .90	− .71	−48.99
II, I, V, VI	I, V, VI	33	313	+ .23	+ .25	+ .48	+15.84
II, I, V, VIII	I, V, VIII	24	80	−.10	− .22	− .32	− 7.68
II, I, V, IX	I, V, IX	81	214	−.06	− .40	− .46	−37.26
II, I, V, VI, IX	I, V, VI, IX	23	129	+1.04	− .70	+ .34	+ 7.82
II, I, VI	I, VI	37	68	+ .01	−1.38	−1.37	−50.69
II, I, VI, VIII	I, VI, VIII	20	60	+ .19	− .10	+ .09	+ 1.80
II, I, VI, IX	I, VI, IX	38	100	+ .11	− .10	+ .01	.38
II, I, VIII	I, VIII	70	155	+ .02	− .20	− .18	−12.60
II, I, VIII, VIII	I, VIII, VIII	52	126	−.28	− .45	− .73	−37.96
II, I, IX	I, IX	83	128	−.28	+ .82	+ .54	+44.82
							−170 + 753 = −.23

The uncorrected weighted average difference in form of II, which was +0.50, would therefore with the connections probably become +0.27.

ing to the method of weighting). The imperfection of our measurements is such that the 1922 and 1923 scores correlate .82 in the case of 1092 boys taken at random.

We may reduce the effect of the error of measurement by relating the gains to the sum of the 1922 and 1923 scores. Doing this also equalizes the negative distortion when gain $(y-x)$ is related to initial score (x) by the positive distortion when it is related to final score (y). The results for 30 large groups of pupils are shown in Table XIV, column B. The median correlation is now +.09 (.08½ or .10 according to the method of weighting). The median difference between the correlation with initial score and the correlation with the sum of the scores is .25 (.24 to .26 according to the system of weighting). According to these determinations, the true correlation between initial ability and gain for pupils taking the same courses, may then be expected to be +.09.

This use of the sum of the 1922 and 1923 scores is open to the objection that the 1923 score is influenced by the gains and so by the studies taken. The resulting error is so slight in comparison with the distortion for which we are correcting as to be negligible for our purpose.

TABLE XVI.—THE DIFFERENCE IN GAIN BETWEEN A PUPIL TAKING A GIVEN SUBJECT AND A PUPIL OF THE SAME SEX AND INITIAL ABILITY WHO TAKES I OR NOTHING IN PLACE OF IT

Subject	Corrected weighted average difference	Sum of weights used in the first determination	Sum of weights used in the correction	Estimated P.E. of the difference
II	+ .27	1191	723	.67
III	− .90	1375	942	.63
IV	+2.92	1026	670	.72
V	+2.33	1720	1305	.56
VI	+1.64	2045	1564	.51
VIII	− .47	2153	1549	.50
IX	+2.64	1261	963	.65
D	− .29	466	149	1.08
T	+ .66	1206	779	.67
M	(too few cases to justify correcting)			

Using +.09 as the true correlation between initial ability and gain, we have, as the regression, gain $= \frac{.09 \text{ (Variability of gain)}}{\text{Variability of initial}}$. This

gives .05 points of gain to be allowed for each point of initial ability.[1] So a group whose 1922 scores averaged 200 would be expected to gain .50 of a point more than a group of the same sex whose 1922 scores averaged 190, if they took identical programs.

That the .09 is a conservative estimate is evidenced by the fact that in a group of 350 colored pupils of mixed programs (not included in the 1637 pupils above) the r for 1922 score with gain was $-.06$, while that for the sum of 1922 and 1923 gain was $+.30$, a difference of .36, though the correlation of 1922 and 1923 was .90. In another group of 1092 boys, some of whom are included in the correlations of Table XIV, the correlation between 1922 and 1923 was .863; that between 1922 and gain $-.135$; and that between the sum of 1922 and 1923 and gain $+.18$; a difference of .31½. In a group of 880 girls, the correlation between 1922 and 1923 was about .85; that between 1922 and gain was .06; and that between the sum of 1922 and 1923 and gain was .22, a difference of 28. If we apply the average of .36, .31½ and .28 to our $-.20$ we have $+.12$.

Turning now to the other method, we have the colored pupils with a central initial score of 117, making a median gain of 13.5, whereas the white pupils of the same city with a central initial score of 192, make a median gain of 23.5. Here a difference of 75 points in initial score corresponds to a difference of 10 points in gain, whereby 1 point of initial score needs an allowance of .133 points in gain.

We have then from independent evidence, two estimates of the allowance for differences in initial ability, .050 per point, and .133 per point, of which the .05 is by the direct use of groups, each composed of pupils of the same sex taking the same combination of studies.[2]

We may therefore correct conservatively for initial inequalities by adding 0.05 point of gain for each point of initial score below 180,

[1] We use the ratio of the obtained variabilities for a total group of 1092 boys (41.8 for initial score and 22.9 for gain). This is only approximate for several reasons. An exact and adequate procedure would require the computation of separate regressions for each of the 30 groups of Table XVI, and the consideration of the effect of the chance error on the variability of the 1922 score and on the variability of the gain. The errors involved in our procedure are not large, and, if our reasoning is correct, act on the whole to make our allowance for initial score too low, and so to make our final estimate of the differences in favor of Physical Science, Mathematics, and Languages, over Biological Sciences, Social Sciences, and Handwork, too favorable to the former. This leaning in the conservative direction is perhaps not undesirable, though we shall, of course, provide the fuller treatment whenever we have the time and facilities.

[2] The same, that is, after the grouping already described.

and subtracting 0.05 point of gain for each point of initial score above 180. A rough but impartial equivalent of this procedure is to do so for only the cases involved in the main differences of Table XI and the other similar tables whence Table XII is derived, defining "main difference" as those where the smaller n is 20 or over. The minor differences may be assumed to follow the same law. This is what we have done. When this correction and that for sex are made in all the tables whence Table XII is derived, we have Table XVI which shows the effect of taking any study rather than I or nothing, if sex and initial ability had both been constant.

The procedure is shown in Table XV wherein the differences of Table XI ($g_1 - g_2$) are corrected first for sex and second for initial ability. Tables similar to Table XV but for Studies III, IV, V, VI, VIII, IX, Physical Training, Dramatic Art, and Manual Training were made. The grand total results for all of these are shown in Table XVI, which gives the effect which taking a certain study for a year rather than taking a "I" study or none, would probably have if half of those taking it were boys and half girls, and were the same as the general average in initial ability.

These corrections reduce the differences between the more intellectual subjects like V, VI, and IX and the more manual subjects of VIII markedly, since the VII's are taken much more often by girls than by boys, and somewhat more often by dull pupils than by bright pupils.

Table XVI presents the best estimate of the influence of a year's study of the subject in question upon ability to score well in standard tests of intelligence that we can derive from our 8564 cases without an enormous further expenditure of time. The probable errors of its determinations can not be computed precisely within the time at our disposal, involving as they do differences of city, grade, age, sex, initial ability, and sampling in relation to studies, and the other factors introduced by our combinations of studies into study groups. To obtain the estimated probable errors of Table XVI, I have simply taken 1½ times the mere individual sampling error of the first determination, counting the mean square deviation of the gains as 22.9, which is the case for the entire group.

The facts of Table XVI may be made more striking and retainable by considering some cases based on them. Suppose, for example, that 3000 pupils beginning the second year of high school were divided into three groups of identical composition, each taking four courses, and that these were for one group, Latin, French, Algebra, and Geom-

etry; for the second, Arithmetic, Bookkeeping, Stenography, and Typewriting; for the third, Cooking, Sewing, Physical Training, and Dramatic Art; for the fourth, three courses in Physical Science and one course in Mathematics. According to Table XVI, the first group should show a gain of 31 score points; the second, a gain of 29 score points; the third, a gain of 22½ score points; the fourth a gain of 33½ score points. How much one of these thousands of pupils would gain in intellectual power, if, instead of taking Cooking, Sewing, Physical Training, and Dramatic Art, they left school and worked at a trade or in a business in such ways as young people of this level of ability would take, our data do not inform us. Common sense, however, would surely place the gain well above zero; the consensus of 10 psychologists and experts in high school education to whom the case was presented, placed it at 18. Between the most favorable and the least favorable programs that we can derive from Table XVI (Bookkeeping, Arithmetic, Physics, and Chemistry against Cooking, Sewing, and two courses in Biology[1]), the spread would be from 34 to 20¾. This is less than the spread between very intellectual boys and average boys[2] of 16 taking the same studies.

A still more concrete illustration may impress the facts even more emphatically. Colleges, so far as I know, universally refuse to credit Stenography and Typewriting for admission, no matter how intellectually superior the candidate is. If some serious-minded boy in the top 5 per cent of his grade wished to take Stenography and Typewriting instead of two years of Latin, claiming that the former would help him better in the life and work of a scholar or man of science, the colleges would reply that the practical gain to him would be 10 or 100 times offset by the loss in intellectual training—in power to think. Just how much would that loss be? By our best available present knowledge, it would be 4 points of gain, or about what the boy could balance by taking four courses in Physical Science instead of four courses in Latin, or by taking two courses in Physical Science or Mathematics instead of two courses in the Social Studies. Moreover, boys of that degree of ability would actually gain much more in ability to think during a year, if they studied X, Y, Z, and Stenography and Typewriting, than the average Grade X pupil would gain during two years if he studied X, Y, Z and Latin.

[1] I do not use all Arithmetic and Bookkeeping or all Biology, since our data do not include cases of such specialization.

[2] Average of *all* boys of 16 not of those who continue in school to Grade X.

Let us now consider the results if we deliberately take the most favorable possibility for large differences between studies, by assuming that the gain from the experience of the examination itself is a special practice effect which is due wholly to that experience and is just as great for one program as for another, and has nothing to do with the ability of the mind to think, or is correlated with it negatively, so that 11.9 points should be subtracted from all these estimates. We then have as the effect of the different programs:

22½ for the 3 courses in Science and 1 in Mathematics

19 for 1 course each in Latin, French, Algebra, and Geometry

17 for 1 course each in Arithmetic, Bookkeeping, Stenography and Typewriting

10½ for 1 course each in Cooking, Sewing, Dramatic Art, and Physical Education

20½ for the best 1 per cent in initial ability[1]

1½ for the lowest 1 per cent in initial ability[1]

11 for the average white pupil

1½ for the average colored pupil

The gain from the more intellectual studies is larger relatively to that from Cooking, Sewing, etc., than it was before, but so also is the gain due to greater initial ability, or to being a white rather than a colored pupil.

The results of Table XVI are in pronounced opposition to the traditional view that certain subjects produce much more general improvement in ability to think than others, and that among the subjects taught in high schools, languages and mathematics are the two that do this to the greatest degree.

If we take, on the one hand, our Groups V (Algebra or Geometry or Trigonometry) and VI (Latin and French, which it will be remembered made the better showing of the two), and, on the other, our Group VIII (Cooking, Sewing, and Stenography), the best estimate of the greater gain in ability to think, due to spending a quarter of one's time during a high school year upon the former rather than the latter, is a little under 2½ points. This is only a quarter of the greater gain that one makes during a year by being white rather than colored, is less than the greater gain that one makes by being a high school boy rather than a high school girl in these cities, the subjects taken being identical! We find notable differences in gain in ability to think as measured

[1] These are approximate estimates, using white pupils in Grades IX, X, and XI in city high schools.

by these tests, but they do not seem to be due to what one studies. Whites gain 23.5 compared with 13.5 for the colored pupils in the same city. The pupils who make scores of 475 to 575 in the sum of the 1922 and 1923 tests, gain about 13 points more than pupils who make scores of 175 to 275. When, however, we measure the effects of differences in studies, race and sex and initial ability of the students being equalized, the effects are small and uncertain.

If our inquiry had been carried out by a psychologist from Mars, who knew nothing of theories of mental discipline, and simply tried to answer the question, "What are the amounts of influence of sex, race, age, amount of ability, and studies taken, upon the gain made during the year in power to think, or intellect, or whatever our stock intelligence tests measure," he might even dismiss "studies taken" with the comment, "The differences are so small and the unreliabilities are relatively so large that this factor seems unimportant." The one causal factor which he would be sure was at work would be the intellect already existent. Those who have the most to begin with gain the most during the year. Whatever studies they take will seem to produce large gains in intellect.

The effects are also qualitatively not what the traditional theories would expect. The order of Table XVI is: Arithmetic and Bookkeeping, first; Physical Science, second; Algebra and Geometry, third; Latin and French, fourth; Physical Training, fifth; the group composed of Civics, Economics, Psychology and Sociology is sixth; our miscellaneous group of History, Music, Shop, Spanish, English, Drawing, and Business (the I's) is seventh; Dramatic Art, Cooking, Sewing, and Stenography, and the Biological Sciences come last. This order seems less satisfactorily explained by the traditional theory than by a mixture of chance, the limitations of our tests, special training, and general training.

For example, the superiority of Mathematics to Latin and French may be a chance result of the small numbers of individuals concerned; there is about one chance in four that with 85,000 instead of 8500 pupils, Mathematics would fall below them. The superiority of Arithmetic and Bookkeeping may be due to the special training which they give in Arithmetic. The low position of the Biological Sciences may be due to the fact that the examinations did not test intellect with any three-dimensional objects or with any acts of living things.

We had hoped to compare the subjects, using the linguistic, mathematical and spatial tests separately, and thus ascertain whether what

we have so far treated as a general improvement in intellect may really be in part specialized. Our 8500 pupils are, however, not enough for this to be done at all precisely. This part of our plan can be carried out adequately only by the addition of 10,000 or 12,000 more pupils' records. Until it is carried out, we can not separate the general from the special effects of the study groups.

Our treatment of the measurements is incomplete in two other respects. We have not equalized for grade or for age.

There are appreciable differences between the gains in relation to grade. Pupils in Grade X in 1923 showed a median gain of 22.5; those in Grade XI in 1923 showed a median gain of 23.8; those in Grade XII in 1923 showed a median gain of 22.4. We do not, however, correct for this, chiefly because the differences in gain according to grade may be due to the differences in studies, which are just what we are trying to isolate. Algebra and Geometry occur predominantly in Grades X and XI. Cooking, Sewing, and Stenography occur predominantly in Grades XI and XII. If we had three times as many cases as we do have, we might make all our comparisons separately for each grade. As things are, the results would not justify the expense of labor, which would be better spent in extending the experiment itself.

The effect of age, grade being kept constant, might well be studied, though its chief effect has been already allowed for by our treatment of gain in relation to initial ability. The younger pupils in a grade will be the abler, and, for our ages will gain more, chiefly because they are the abler. As with grade, we need many more cases than the 8564 we have, in order to free our results from age insofar as it acts otherwise than by its relation to initial ability.

It should be noted that the facts of Table XVI are not in opposition to the theory that general improvement of intellect results from school studies, and that studies vary in their value in this respect. The question is one of quantity. The traditional theory was (1) that the amount of general improvement due to studies was large; and (2) that the differences between studies in respect of it were large, so that the value of a study as training for the intellect was of very great importance relative to its value as special training in mathematical or linguistic or spatial or other thinking; and that (3) mathematics and languages gave much more of this general improvement than other studies did. The facts of Table XVI, if corroborated by similar experiments, prove that the amount of general improvement due to

studies is small; that the differences between studies in respect of it are small, so that the values of studies may be decided largely by consideration of the special training which they give; and that the languages have no claims to preeminence.

It may be noted further that our experiment does not, except perhaps indirectly, measure the effect of studies upon ability to put forth effort, or to stand the strain of discomfort, or any others of the alleged disciplinary values of studies, save their effect upon ability to think as measured by accepted tests of "intelligence." If the memories of individuals concerning the general training which they themselves received from different studies in their own school days, can be trusted, the facts for the total effects of different high school studies are much the same as our experiment has disclosed in the case of their effects on power to think. The facts concerning some hundreds of these memories are presented later in a separate section.

A repetition of this experiment with 16,000 or 18,000 more cases is needed before final conclusions should be stated. Also, certain accessory investigations would be highly desirable. Each reader, however, will doubtless form some opinion of what these final conclusions will probably be, and the writer may be permitted to express the opinion which he himself has formed. It seems to me likely that a considerable fraction of the gain found here is due to the general growth of a year under the stimulation of such a life as a Grade X pupil would lead, in school or out, and to the reaction of his mind to the first and second experiences with the examination. This amount might be set at 18 + 2. Another fraction of it is due to training in using words, numbers, other symbols, and in dealing with ideas. This fraction varies according to the subjects taken, from somewhere near zero in the case of a program made up of Physical Training, certain forms of Sewing and Typing and the like, up to 8 or 10 in the case of a program made up of an abstract science like Physics, systematic Grammar, Algebra, and the like. Whether this training is called special or general will depend upon a person's terminology. The training in Algebra is general in the sense that it spreads beyond mere Algebra to influence probably several of the tests included in our composite. It is specialized in the sense that probably a test concerned with thought about problems in politics, philanthropy, and literary criticism, might show very slight or zero improvement due to it. There may remain a small balance of two or three points of gain out of 30 in the case of a pupil taking such a program of abstract intellec-

tual studies, which is due to a still more general improvement of thinking for any sort of data; but this is doubtful.

By any reasonable interpretation of the results, it appears that the physical sciences are equal, if not superior, to languages and mathematics as now taught in our American cities, in respect of mental discipline from any point of view. The social studies, including History and English would probably make a better showing than they do if the examination tested thinking with respect to broad questions of human nature, morals, and community affairs, instead of some of its linguistic and numerical items. I doubt, however, whether the difference is wholly due to the content of the test. Something must probably be conceded to the more rigid and systematic and abstract nature of the thinking in Mathematics, Grammar, and Physics as compared with that in English and History. As already suggested, the biological sciences probably suffered in a similar manner from the limitations of the tests.

By any reasonable interpretation of the results, the intellectual values of studies should be determined largely by the special information, habits, interests, attitudes, and ideals which they demonstrably produce. The expectation of any large differences in general improvement of the mind from one study rather than another seems doomed to disappointment. The chief reason why good thinkers seem superficially to have been made such by having taken certain school studies, is that good thinkers have taken such studies, becoming better by the inherent tendency of the good to gain more than the poor from any study. When the good thinkers studied Greek and Latin, these studies seemed to make good thinking. Now that the good thinkers study Physics and Trigonometry, these seem to make good thinkers. If the abler pupils should all study Physical Education and Dramatic Art, these subjects would seem to make good thinkers. These were, indeed, a large fraction of the program of studies for the best thinkers the world has produced, the Athenian Greeks. After positive correlation of gain with initial ability is allowed for, the balance in favor of any study is certainly not large. Disciplinary values may be real and deserve weight in the curriculum, but the weights should be reasonable.

A STUDY OF EXTREME CASES OF UNRELIABILITY

PERCIVAL M. SYMONDS

University of Hawaii

In order to discover, if possible, some of the causes of the unreliability of mental tests I made a study of extreme deviations between the scores on the two forms of the National Intelligence Tests, both Scale A and Scale B. The study of the extremes of any distribution is very fruitful in leading to the suggestion of possible relationships, but is ill adapted to the validating of actual relationships. In order to prove relationships the whole range of the distribution must be studied and whatever correlations there may be must be determined over this whole range. The abnormal psychologists have led the way to the discovery of causes and relationships. But what appears at first sight to be an extremely important relationship in an extreme case may turn out later to be insignificant when considering people in general. Hence any suggestions which follow should be validated by experimentation and correlation.

The National Intelligence Test, Scale A Form 1, was given on a Tuesday to the Territorial Normal Training School. On Wednesday, Scale B Form 1 was given. On Monday of the next week, Scale A Form 2 was given, and on Tuesday, the day following, Scale B Form 2. I made a tabulation of the deviations of scores on each test of Form 2 over Form 1. This study concerns the extremes of these distributions of deviations.

To give an example, following is the distribution of deviations for N.I.T. A_1:

Deviation	
12	3
10	2
8	4
6	16
4	40
2	45
0	65
− 2	32
− 4	14
− 6	4
− 8	4
	232

This table is to be read as follows: "Three children made a score on Form 2 on N.I.T. A Test 1, 12 points higher than on Form 1." In this example I studied the tests in detail of the 9 children who made deviations of +8 or over and of the 8 children who made −6 or lower. I tabulated for each child the score on Test 1 for the two forms, the

n	Deviation	Score on test		Total score		Number tried	Right	Wrong	Number tried	Right	Wrong
		First trial	Second trial	First trial	Second trial	First trial			Second trial		
A_1 9	+ 9.8	6.0	15.8	78.0	96.4	10.1	3.0	7.1	11.0	7.9	3.1
8	− 7.0	15.0	8.0	79.5	84.9	9.4	7.5	1.9	9.0	4.0	5.0
A_2 17	+10.9	12.6	22.4	73.1	88.2	12.6	5.7	6.9	14.4	11.2	3.2
17	− 7.1	20.7	13.7	92.9	92.7	14.6	10.3	4.3	12.5	5.7	6.8
A_3 13	+13.3	16.9	30.2	81.4	103.0	34.6	16.9	17.7	43.8	30.2	13.6
15	− 6.4	22.6	16.3	74.3	69.4	37.8	22.7	15.2	34.1	16.3	17.8
A_4 12	+12.1	11.7	23.8	93.0	114.5	31.2	21.8	9.4	34.8	28.9	5.9
12	−11.9	17.6	5.7	94.7	86.2	29.9	23.8	6.2	35.2	20.4	14.8
A_5 14	+17.3	9.3	26.6	52.3	78.9	38.4	31.2	7.2	88.8	88.3	0.5
12	− 4.4	29.6	24.2	93.8	95.4	95.5	94.8	0.7	81.8	81.1	0.7
B_1 12	+ 8.3	19.2	27.5	84.5	98.7	14.4	9.6	4.8	16.1	13.7	2.3
11	− 7.1	28.0	20.9	81.5	86.5	16.4	14.0	2.3	16.4	10.4	5.9
B_2 10	+ 6.5	7.0	13.0	58.6	75.6	16.0	7.0	9.0	24.9	13.5	11.4
9	− 8.9	22.0	13.1	102.8	101.9	30.9	22.0	8.9	29.0	13.1	15.9
B_3 13	+13.0	9.0	22.0	79.6	98.7	30.5	19.9	10.8	34.7	28.3	6.3
12	− 9.4	19.4	10.0	92.7	83.5	30.2	24.8	5.4	33.9	21.8	12.0
B_4 8	+ 9.4	5.4	14.8	91.5	105.0	20.0	5.4	14.6	26.1	14.7	11.4
13	− 4.7	12.3	7.6	84.6	82.0	27.4	12.3	15.1	25.6	7.6	17.9
B_5 11	+15.0	16.3	31.3	76.2	93.0	30.4	22.6	7.7	35.6	33.4	2.2
7	− 7.4	29.0	21.6	78.5	70.9	32.7	30.8	1.9	26.4	24.0	2.4
Averages for Total Scale											
A+ 65	+12.7	11.4	24.2	75.2	95.1	25.6	16.0	9.6	39.5	34.4	5.2
B+ 54	+10.6	11.8	22.5	77.9	94.1	22.7	13.6	9.1	27.7	21.4	6.3
A− 64	− 7.3	21.4	14.0	87.1	85.5	37.4	31.2	6.2	34.3	25.0	9.5
B− 52	− 7.4	21.2	13.8	87.9	85.2	27.0	19.7	7.3	26.2	14.6	11.6
A+B+	+11.7	11.6	23.4	76.4	94.5	24.3	14.9	9.0	34.2	28.5	5.7
A+B−	− 7.3	21.2	13.9	87.5	85.5	32.8	26.1	6.7	30.7	20.2	10.4
A+−	+ 3.0	16.3	19.1	81.1	86.5	31.5	23.6	7.9	37.0	29.6	7.3
B+−	+ 1.8	16.1	17.9	82.6	89.7	24.8	16.6	8.2	26.9	18.1	8.8

total score on the test for the two forms, the number tried, the number right and the number wrong for both forms. The table shows the results, averaged for each group of plus and minus deviation.

Note first that the average plus deviation exceeds the average minus deviation in every case except B_2. This excess of plus over minus deviations is quite certainly the practice effect coupled with whatever difference in difficulty there happened to be, this practice effect showing itself in the extremes of the distributions as well as in the averages. B_2 is a case where Form 2 is more difficult than Form 1.

When one looks at the total scores on the tests there seems to be a positive correlation between the deviation in a single test and the deviation in total score—a high plus deviation in a single test goes with a high plus deviation in the total test and vice versa. But one must remember that this single test under consideration goes toward making the total deviation for the whole test. If the deviation for the single test is subtracted from the deviation for the total test, the result shows that there is practically no correlation between deviations in the single test and total tests.

	Single Test Deviations	Total Test Deviations	Difference
N.I.T.A.	+12.7	+19.9	+7.1
	− 7.3	− 1.6	+5.7
N.I.T.B.	+10.6	+16.2	+5.6
	− 7.4	− 2.7	+4.7

Correlations have not been computed for these tests, but with other tests correlations of −.07 and +.03 were found between the deviations in one test and another test given immediately following on both occasions. This is evidence that the unreliability of a test is not affected noticeably by the general condition of the individual—state of health, nervousness, listlessness or eagerness, etc., but to factors which are peculiar to each test in itself. What some of these factors are will presently be brought out.

Another method of looking at the same relationship is to determine how many times the same individual will make a high plus or minus deviation in taking N.I.T.A or B.

A+ Number Tests High Plus Deviation	Number Individuals	A− Number Tests High Minus Deviation	Number Individuals
0	174	0	178
1	52	1	45
2	5	2	8
3	1	3	1

B+		B−	
Number Tests High Plus Deviation	Number Individuals	Number Tests High Minus Deviation	Number Individuals
0	191	0	195
1	48	1	43
2	3	2	3
3	0	3	1

8 individuals had both a high and low deviation on A
9 individuals had both a high and low deviation on B

This shows that almost all high deviations on one test are single occurrences on the battery for the individual. Much more rarely does the same individual make another large plus or minus deviation in the same battery. In fact, individuals who make a high plus deviation in one test in a battery are as likely to make a high deviation as another high plus deviation on some other test in the battery.

11 individuals had a high plus deviation on both A and B
14 individuals had a high minus deviation on both A and B

One might as well once for all give up the bogey that frightens all novices in testing that a child's nervousness, or state of health, or enthusiasm decreases the reliability of a test. So many times one runs across the statement even in works on measurement that the tests should be checked up with other evidence because single tests do not give a student opportunity to show his real ability. Courtis says, "The best examination is not that represented by the score of a single performance in a single day. Human effort is variable and human skill too easily upset to make it fair to have promotion based upon chance scores." But the chances are that an individual will not deviate from his real ability on a test and the infrequent thing is for a pupil to be affected greatly by some extraneous circumstance.

Differences in speed is one factor causing the unreliability of a test. Those who made high plus deviations tried 9.9 more items on the average in the second test than the first while those who made high minus deviations tried 2.1 items less. An examination of the figures shows that this difference in speed is largely responsible for the practice effect. As I look at the individual papers this cause for unreliability stands out clearly. In many of the tests a child seems to be struggling with what to do the first time. Especially is this true with tasks that are unfamiliar. Test 3 of A, finding two words that tell what a thing always has, is an example. Here the children often go much faster the second time. At first the problem is not clearly

defined, they are struggling with what to do, and it is not until the second trial that they can go ahead with a clear understanding of the task. Test 5 of A, the substitution test, is another clear example. Test 2 of B, the information test and Test 4, the mixed relations test also show the same phenomenon. Many children, unfamiliar with the task of underlining, or with the mental processes needed, are slow to get started. On the other hand in tests like A_1 and B_1, the arithmetic tests, where the tasks are familiar, or in A_4, the opposites test, where the process is easily comprehended, gain in speed is not such a factor in unreliability.

Why the minus deviations should show a slowing up of the work is more difficult to understand. In some few cases the explanation may be simple listlessness or inertia. But in some instances it was clearly evident that when they finally did catch on and comprehend the problem they "stopped to think." To use a simile, it is very much like the electric car whose wheels are slipping when starting, and then are slowed up by the friction when sand is dropped into the rails. Test A_3 and B_4 give examples of this. In the first trial the child may simply underline indiscriminately, getting a score by chance. But in the second trial he comprehends the problem, and, stopping to think, is retarded.

Accuracy is another factor causing unreliability. The following tabulation shows the percentage right to number tried in each test.

PER CENT CORRECT ON EACH TEST OF THE NATIONAL INTELLIGENCE SCALES

	First trial, per cent	Second trial, per cent		First trial, per cent	Second trial, per cent
A_1+	28	72	B_1+	67	85
−	80	44	−	85	63
A_2+	45	78	B_2+	44	91
−	71	46	−	71	45
A_3+	49	69	B_3+	65	82
−	60	48	−	82	64
A_4+	70	83	B_4+	27	56
−	80	58	−	45	30
A_5+	81	99	B_5+	74	94
−	99	99	−	94	91

In every case a high plus deviation has been accompanied by increase in accuracy, and a high minus deviation has been accompanied by decrease in accuracy. The distressing thing in these figures is the low accuracy per cents. Accuracies of 44, 30 or 27 per cent are too low and show the wrong set that so many children have with regard to school work. Something is required of them. If they move their pencils, make marks on the paper and go through the form of the task they will not be liable to immediate censure and probably will escape censure altogether. Again there is the set that an examination is graded on the percentage basis, that they must get a certain percentage to pass and that the least they can do is to cover the test. It is plainly evident that some children rush blindly through a test of 40 exercises to finish before time is called, the exhilaration of having at least finished being the only satisfaction they get in lieu of knowing their score immediately. But the same factor that we noticed in considering speed is evident here—"catching on" to the problem leads to greater accuracy. The child who underlines more than two words in a line in A_3 or fills in the squares in A_5 in numerical order is bound to be more accurate when he finds out what he is to do. Likewise the child who treats A_2 as a free association test is less accurate than he is when his associations are controlled by the sense required. Perhaps nervousness is a factor in this accuracy. If it is, it is queer that it is a disease that attacks only one test.

In the case of high minus deviations accuracy is also a factor. My notes read as follows: "hurried-careless," "harder-tendency to let go of set," "faster-careless toward end," "less check on self." Here are cases where the set, whether it is opposites, mixed relations, definitions, etc., is let go and the child drops back to indiscriminate underlining, probably in the endeavor to speed up and finish.

Finally, in the case of the tests which are scored right minus wrong there seemed to be no other reason for the deviation except pure chance. That this is plausible has been shown by the recent work of West, Chapman, and others.

In enumerating these causes of unreliability I have touched only one aspect of it. Another aspect is the construction of the test. Of the latter this study has nothing to say.

In conclusion I would like to suggest the possibility of lowering test unreliability by means of systematic training. Children should be taught to take tests. In the first place children need to be taught to go through the motions that modern testing demands. They need to

be taught to put down answers on the dotted line, that they may figure on the sides of the paper if need be, what it means to underline, or cross out, to follow directions if they are told to underline only one word or two words, to write letters where they should be written, etc. To be sure, this is provided for in the five exercises in the National Intelligence Tests, but this does not guarantee that the dull child will learn the process in the half-minute given. Specific training should be given. Again training in the processes themselves should be given. For some children the fore-exercise is not sufficient even here. The substitution test, the mixed relations test, finding two words that tell what a thing always has, are unfamiliar processes to young children and need special training. I suspect that certain children with psychopathic tendencies find it particularly difficult to hold the set after it is obtained.

Then again, definite training needs to be given in the taking of tests themselves. A child needs to learn that getting an example right is the only thing that scores. He needs to be taught that he is not expected to finish the test but to do as much as he can do *well*. He needs to be taught that getting a certain percentage on the test is not what counts, but what he does in comparison with others of the class. I suspect that the best training in this is by showing the children their scores and by showing them where they stand with respect to others in the same class. I also suspect that before this can be accomplished the mental sets which teachers inherited from their own school days will have to be changed first. Merely telling these things in the directions of a test does not teach them to all children.

These suggestions if carried out ought to increase the reliability of tests in the class which has been trained. It ought also to make intelligence tests better measures of intelligence. While this training would eliminate those factors of intelligence in taking a test, such as greater initial adaptability, ability to follow directions, and the like, it would make the tests truer indications of ability in the sense of "how hard a thing a person can do."

Conclusions

1. The general conditions of an individual—nervousness, general health, enthusiasm, etc., exert practically no influence under ordinary school conditions on the results of testing.

2. Individuals who do extremely poorly on a certain test in relation to their true ability are quite as likely to do extremely well on another

test immediately following as they are to do poorly again, and *vice versa*.

3. Differences in the speed of work in a test is a cause for the unreliability of the test.

4. Differences in accuracy in a test is a cause for the unreliability of the test.

5. The causes for these differences in speed and accuracy are surmised to be slowness in understanding the mechanical operations of the test, failure to comprehend the mental processes required, failure to hold this "set" throughout the test, and failure to have the proper attitude toward taking a test.

6. Reliability ought to be increased by training directly toward correcting these faults.

PART AND WHOLE METHODS OF LEARNING

H. B. REED

Grinnell College, Grinnell, Iowa

Practically all writers of psychological textbooks favor the whole method of learning. E. Meumann, who was one of the first to discuss this method at length said, "We may therefore regard the following general result as established: For adults and children it is more advantageous and it is psychologically and pedagogically more appropriate to learn every sort of material as a whole than to break it up into parts." ("Psychology of Learning," p. 249.) A. I. Gates, in a book which came off the press as recently as last May, says, "The whole method shows to advantage in memorizing poetry, prose, vocabularies, lists of words, and non-sense syllables, and probably in most materials which are to be memorized in a serial order." (Psychology for Students of Education, p. 292.) Statements in favor of the whole method may also be found in the texts of Colvin, Freeman, Pillsbury, Norsworthy and Whitley, Starch, Hunter, Woodworth, and others. Just why psychologists have fallen in love with the whole method is difficult to say, but an analysis of the evidence in published experiments upon this problem shows that the facts are just as much, if not more, in favor of some form of the part method. Meumann gives us some twenty-five pages of argument in favor of the whole method, in spite of the fact that at the time of his writing only one experiment, that of Steffens, was in favor of the whole method. The facts discovered in Meumann's own laboratory as published in the experiments of Ebert and Meumann, and of Pentschew, favored the part method insofar as economy is measured by learning time. The facts discovered by Ephrussi also favor the part method for vocabularies and nonsense materials. Meumann admits this but says that the results are not pedagogically significant. The results discovered in his own laboratory can be interpreted only in favor of the whole method if economy is measured by number of repetitions rather than by learning time and even then the evidence is conflicting. Of the more recent experiments those of Lakenan, and of Pyle and Snyder, favor the whole method while those of Pechstein favor the part method. If the evidence in favor of each method is impartially weighed, as every scientific writer should do, we must judge the statements quoted above from textbooks based on argument and generosity rather than on scientific evidence. At best the results in favor of either method were obtained from very

few individuals, far too few to establish anything general in psychology. We shall see a bit later that the most economical method of learning is an individual trait rather than a general psychological law. My own experiments indicate that for most individuals a part method is the more economical.

To justify some of the above statements, I will reproduce some of the findings of Ebert and Meumann, and of Pentschew.

Ebert and Meumann had five subjects learn two rows of twelve nonsense syllables each every day for thirty-two days. Four methods were followed: *A*, the whole; *B*, the part (learn first half; then second half, then two halves together); *C*, mixed (pause one second at the end of the fourth and eighth syllables respectively); and *D*, mixed (pause two seconds at the end of the sixth syllable). Table I gives the results:

TABLE I

METHOD	*A*	*B*	*C*	*D*
Average number of repetitions to learn	10.5	9.2	7.9	9.4
Average number repetitions to re-learn	4.4	4.7	4.1	4.4

I repeated this experiment with the following modifications: I used eight subjects. Two learned nonsense syllables 60 minutes, and the other thirty minutes a day for 15 days. Economy was measured by time. Table II gives the results.

TABLE II

METHOD	*A*	*B*	*C*	*D*
Average learning time	7.25	6.31	5.83	6.80
Average re-learning time	3.12	3.32	2.92	2.49
Size of unit in which subjects learned	12	6	4	6
Average memory span of all subjects		4.45		

In both experiments the whole method is the longest in learning and the mixed methods are the shortest. Methods *B*, *C*, *D*, are forms of the part procedure. Many reasons have been given for the alleged superiority of the whole method but the results suggest that the fundamental factor which makes the part procedure economical is that it adjusts the material to the memory span of the learner. These facts were shown by the author six years ago but they seem to have been unnoticed by the writers of textbooks who have persisted in preaching the superiority of the whole method. (See Repetition of Ebert and Meumann's Experiment on Memory, *Jour. of Exp. Psy.*, Vol. II, 1917.)

The following table is a sample of the findings of Pentschew.

TABLE III

	Repetitions	Time Minutes	Time Seconds
3 children, one stanza in two parts	17.2	3	17
3 children, one stanza as a whole			
Visual	13.3	3	54
Aloud	18.2	5	16
One adult, 3 stanzas in four parts	33.5	24	17
One adult, 3 stanzas as a whole	17.5	36	24
3 children one 12 syllable row in three parts	24.8		
3 children one 12 syllable row as a whole	28.2		
5 adults one 12 syllable row in three parts	18.		
5 adults one 12 syllable row as a whole			
Visual	21.3		
Aloud	16.5		

It will be seen that as measured by time both children and adults find the part method the more economical. As measured by repetitions children find the part method the more economical for nonsense syllables. The same is true for adults if the presentation is visual, but not if it is oral. In the face of these facts both Meumann and Pentschew multiply their objections to the part method and give long arguments, the conclusion of which is that it is established that the whole method is both psychologically and pedagogically the better for all material.

Individual Experiments with Poetry and Prose

I shall now describe the results of my own experiments. These may be divided into three groups: (1) four experiments with single individuals upon prose and poetry; (2) group experiments showing the influence of the part and whole methods of reading upon the comprehension of prose; (3) experiments with large number of individuals showing the influence of method on the learning of poetry.

First Group

1. Subject *A*, college sophmore, boy, memorized Arnold's essay on Numbers. The odd paragraphs were memorized by the part progressive method: he first memorized the first sentence, then the second sentence then the first two sentences together, then the third sentence, then the first three sentences together, etc. The even numbered paragraphs were learned by the whole method.

TABLE IV

Method	Number of Lines	Average Number of Seconds per Line
Progressive part	218	45
Whole	214	60

2. Subject *B*, college graduate, teacher, memorized stanzas from Byron's poem, *Childe Harold.* Alternate stanzas were committed by the progressive part method, the stanza being divided into four parts. The other stanzas were learned by the whole method.

TABLE V

METHOD	NUMBER OF STANZAS	MINUTES PER STANZA
Progressive part	10	9.10
Whole	10	11.17

3. Subject *H*, girl, college sophmore, memorized Wordsworth's poem, *Michael.* Alternate stanzas by the part and the others by the whole method. In the former, the unit consisted of the amount of material between two successive semicolons. This poem was selected because the lines are closely knit together, and the stanzas were purposely broken into parts that would destroy the meaning.

TABLE VI

METHOD	NUMBER OF LINES	AVERAGE NUMBER OF SECONDS PER LINE
Part	243	73
Whole	229	51

4. Subject *D*, boy, college sophmore, learned stanzas from Pope's poem, *Annus Mirabilis.* He learned four groups of stanzas, one half of the stanzas of each group were learned by the whole method and the other half by the part method. In Group I, the unit for the part method was the couplet, but in the remaining groups it was the stanza. The stanzas were first learned separately and then together. Each set of stanzas was re-learned after 24 hours by the same method used in learning. A special feature in this experiment was that the merit of each method was tested in relation to varying quantities of material. For example, in Group 1 the methods were compared in learning one stanza at a time; in Group 2 at learning two stanzas at a time; in Group 3 in learning four stanzas at a time and in Group 5 in learning six stanzas at a time. Table VII gives the results.

Here it is to be observed that in every case, the part method is the more economical both in learning and in re-learning, and that this is true for large as well as small units.

All the experiments of this group point to the general superiority of the part method in learning meaningful material whether poetry or prose. Experiment 3, however, suggests that the part should be a meaningful unit, such as a sentence or a stanza. The results

TABLE VII

Group	Number of stanzas	Method	Average seconds per unit to learn	Average seconds per unit to re-learn
1	6	1 stanza in 2 parts	96	38.8
		1 stanza as a whole	100	39.9
2	12	2 stanzas in 2 parts	242.5	106.3
		2 stanzas as a whole	317	102.5
3	12	3 stanzas in 3 parts	370.5	178
		3 stanzas as a whole	509	186.4
4	16	4 stanzas in 4 parts	516.5	283.8
		4 stanzas as a whole	558	333
5	24	6 stanzas in 6 parts	906.6	427
		6 stanzas as a whole	1010.5	497.5

of these four experiments as well as those obtained from the repetition of the work of Ebert and Meumann should be balanced against the hitherto published experiments which were made on a small number of subjects. If they are, it is doubtful if any general statement can be made in favor of the whole method.

Second Group

The psychology of individual differences teaches us that psychological laws can not be established from the results obtained from a few subjects. Hence no great weight can be attached to hitherto published experiments upon this problem, and the results obtained from the experiments in the first group must be thought of as only suggestive. Mindful of this I began experiments with greater numbers.

The principal argument made in behalf of the whole method is that it gives the meaning so much better than the part, and that all associations are established in the direction effective for reproduction. If this is true the whole method of reading should show a particular advantage in the study of prose, such as a chapter on history or psychology. In this connection it should, however, be pointed out that in most of these selections the thought of a given chapter is usually

developed in a logical order so that the comprehension of the later parts is dependent upon the comprehension of the earlier parts. If, then, the thought of a paragraph is not obtained in the first reading, a rereading will be profitable not only for that paragraph but also for those that come later. That is, a part procedure would seem in such cases to be well adapted to the comprehension of meaning. Again, the alleged disadvantage for the part procedure would be only temporary, for eventually it covers the whole piece, and when it does it should give the meaning of the whole. But as a matter of fact all well-written prose is meaningful as it proceeds, and few authors expect a reader not to understand the first or middle of an essay until he has finished the last line. *A priori* reasoning is just as favorable to the part as to the whole procedure. Now, let us look at the facts.

In order to discover the influence of method upon comprehension I had four sections in psychology containing a total of 169 students read Chapters VIII, IX, and X of Freeman's "How Children Learn" by either of two methods, part or whole. In the part, the student read each paragraph twice before going to the next; in the whole, the student read the entire chapter from beginning to end twice in succession. Each section was divided into two groups, Group *A* following the part and Group *B* the whole procedure. In Chapter IX, these groups changed methods, and again when they read Chapter X. The Army Alpha tests were given to two of the sections so that we could check the influence of method against ability. Comprehension was measured by an essay examination on each chapter. The papers were graded on the percentile basis. Each student also kept an exact record of the amount of time spent in reading. The following table gives the results:

TABLE VIII

Method	Number of sections	Average grade	Average Alpha score	Average minutes to read
Part	4	79.94	.....	69.75
Whole	4	80.49	.....	68.40
Part	2	78.25	126.4	
Whole	2	77.39	123.9	

This shows that the two methods are practically equal, and that the small difference in favor of the part method in two sections is due

to a small advantage in intelligence. The factor of intelligence is brought out more clearly in the detailed results of Section 1 as shown in the following table.

TABLE IX

Group	Number	Chapter	Method	Grade	Average Alpha score
A	23	8	Whole	70.7	121.5
B	18	8	Part	77.8	130.7
B	21	9	Part	65.9	122.5
A	16	9	Whole	73.04	124.5
A	15	10	Whole	82.1	126.2
B	19	10	Part	82.4.	125.5

Here it is seen that the two methods approach each other in equality as the two groups become equally balanced in intelligence. A closer analysis however, shows that the advantage of one of the methods depends upon the individual student. Table X shows this for one section of 48 students.

TABLE X

30 found the part method superior.
18 found the whole method superior.
27 found less than 6 points difference.
17 found less than 12 points difference.
4 found less than 24 points difference.

The general superiority of the part method may be due to the former study habits of the students or it may be due to some inherent advantage which the part method has in the learning process of the individual. It would be difficult to find out which of these statements is true.

Third Group

This group of experiments was made upon seven sections in experimental psychology which contained a total of 226 students. The experiment was not made until after the students had become trained to do psychological experiments and to follow directions exactly. The material used consisted of a selection of 12 stanzas of poetry which was divided into three sections of four stanzas each. The procedure was to divide each class into pairs, one member of each pair to act as experimenter and the other as subject. The subject learned the material according to directions while the experimenter kept time and checked up the correctness of the reproduction. Three methods of

learning were tested; part, whole, and part progressive. In the part, the four stanzas were learned separately and then together. In the whole, the four stanzas were learned together as a unit by reading them through from beginning to end until learned. In the part progressive, the subject learned the first stanza, then the second, then the first two together, then the third, then the first three together, etc. The materials were considered learned when the first errorless reproduction was reached. In order to equalize the influence of the factors of adaptation, fatigue and the possible unequal difficulty of three sections of poetry, one-third of the subjects used the three methods in the order P.—P.P.—W.; one-third in the order P.P.—W.—P.; and one-third in the order W.—P.—P.P. Thus each section was learned equally often by each method. The average time in minutes for each method is told in the following table.

Number of Students	Part	Part Progressive	Whole
113	5.45	5.21	5.95

These results confirm the experiments made with smaller numbers repeated above. The *whole* method is the longest in learning, the *part progressive* the shortest, and the *part* between the other two. When we analyze the results according to the individuals we find that out of 113 students

31 find the part method superior.
56 find the part progressive method superior.
26 find the whole method superior.

Again it seems that the advantage of a particular method turns upon the individual learner, although in the average the part progressive is nearly 13 per cent shorter than the whole.

Our experiment thus establishes the result that the whole, instead of being the most economical method of learning as the textbooks tell us, is in fact the least economical for most persons. The same is apparently true for re-learning, but as only two of our experiments measured retention this matter is still an open question.

The explanation of these results is not an easy one, but there must be some inherent advantage in the part method. I am inclined to think that the fundamental reason for the superiority of the part method is that it adjusts the material more adequately to the learner's span of attention. When we come to think of it, the only way in which the mind learns is by parts. We can follow the whole procedure, but after all we memorize by parts. In one reading, this part comes above the memory threshold, and in another reading that part, and so on until the entire piece is memorized. Since this is the way the mind

actually learns, why not follow some form of the part method, and adjust the method to the nature of the mind?

There is really only one theoretical objection to the part method, and that is that it introduces some conflicting associations. But there are many objections to a strict adherence to the whole method. It delays the perception of progress, creates confusion by overstepping the learner's span of attention and wastes time by causing a useless repetition of the known in order to master the unknown. However a scientific writer should not base his conclusions upon theoretical considerations but upon the facts. If this is done we believe that an impartial and careful study of hitherto published experiments will show most of the evidence in favor of some form of the part method.

When to this is added the evidence from the large numbers of learners whose results are presented above we believe that scientific procedure requires us to say that the part method of learning is the more economical for most persons.

SUMMARY

Although nearly all discussions in psychological literature of economical methods of learning favor the whole method, yet a careful study of facts shows a majority of the experiments to be in favor of the part method. This is true even of some experiments the conclusions of which are in favor of the whole method, namely, those of Ebert and Meumann, and of Pentschew.

A repetition of the work of Ebert and Meumann found the evidence in favor of a part method, just as did the original. Out of four experiments with individuals only one favored the whole method.

An experiment conducted with 169 college students showed that the part and whole procedure had on the average equal value for comprehending the meaning of prose, although the study of the results by individuals shows that most of them find the part procedure superior.

An experiment conducted with 113 college students in memorizing poetry showed that the part progressive method was the shortest, that the whole method was the longest, and that the part method was between the other two.

The reason for the superiority of the part method is that it adjusts the material more adequately to the learner's span of attention.

The most economical method of learning is in the last analysis a question for the individual learner.

Most individuals find the part method superior.

A NOTE ON THE SHORTENING OF EXAMINATIONS

DONALD A. LAIRD

University of Wyoming

As has been pointed out in a previous article[1] the chief interest in objective examinations in colleges and universities seems to be in their utility as a labor saving device. In comparison with this all other factors of advantage inherent in a test of this sort are reduced to a minimum if the literature on their use in higher education is to be taken as the sole indication.

In this brief communication data will be presented that indicate clearly several ways in which the examination labor on the part of the instructor can be materially lessened.

1. *A Comparison of a Long and a Short Objective Examination.*—Sixty-seven students in an elementary course in psychology were given an objective test at the close of the fall term. This test comprised eighty subjects that could be answered by a single word or short phrase. None of the replies could be equivocal. The time consumed by the examination was a little more than an hour and a half. The papers were corrected and each paper was divided into four sections, Section *A* comprised the first twenty questions, Section *B* the first forty questions, Section *C* the first sixty questions, and Section *D* the entire paper. Each section was scored on the percentage of correct replies. Then the percentage right in each of the abbreviated sections (*A*,*B*,*C*) was correlated with the percentage right in the full length examination. The product-moment correlations were as follows:

20 questions with 80 questions	+.84
40 questions with 80 questions	+.88
60 questions with 80 questions	+.92

So far as accuracy is concerned it is apparent that a twenty-question objective examination of this type is as reliable, as examinations go, as an examination four times as long. Furthermore, it is a question if the correlation increases enough with the addition of each twenty questions to warrant the additional expenditure of time. Grading accuracy alone should be satisfied with the shorter examination. For the instructor, however, who delights in instilling due respect and fear for examinations a shorter examination would not satisfy.

2. *A Comparison of Students' Estimates of Their Abilities in a Subject with Their Abilities as Determined by Objective Tests.*—As a further

[1] This journal, February, 1923, page 123ff.

probing into time saving in examinations a study was made of students' estimates of the grade they would receive with what they actually did receive in the eighty question objective test described under (1). The same students were used. Before beginning the examination they were instructed to inventory their abilities and write at the head of their paper the grade that they thought they merited in the course. After the examination had been taken they were told to look back over what they had just done and grade themselves again, this time writing the grade at the end of the paper. The students were already familiar with the system of grading in use, in which 5 per cent of the students were given failures and 5 per cent I, 20 per cent II and 20 per cent IV, and 50 per cent the average grade of III. A distribution curve, on the bases indicated, was made of the scores and marks were assigned accordingly. In determining these grades only the showing on the objective test was taken into consideration. "Conditions" and "incomplete" grades were used to take the other factors into consideration.

The product-moment correlation between the estimate before examination and the grade received was +.42. The correlation between the second estimate and the grade earned was +.37. These correlations are not high enough to warrant the substitution of students' estimates for an objective test when all factors are considered, but still they closely approach the correlations that have been found between essay and objective examinations.

It should be remarked in passing that the cause for the lowering of these correlations evidently lies in the fact that students who did best in the examination under-estimated their abilities. There is also a *slight* tendency for students who made failures to over-estimate their abilities. By and large, however, it is the underestimation of abilities by the better students that lowers the correlation.

The reason for the lowering of the correlation in the estimate after the examination has been taken probably lies in a general under-estimation. Judging from the estimated marks most of the students feel certain that they have done much less than they really did. In the first estimate the central tendency of the marks is near the average mark of III. After the examination had been taken there is a lowering of the central tendency by a little more than half a division.

NOTE CONCERNING A T-SCALE FOR WILSON LANGUAGE ERROR TESTS

G. M. WILSON

Boston University

On the basis of the returns of 485 twelve-year-olds in Duluth, Minnesota, a graduate group in education recently figured out a T-scale for the language error tests. Those familar with Dr. McCall on "How to Measure in Education" will understand the method and the value of a T-scale. No twelve year old pupil has scored lower then two correct answers, so the scale begins at 2, continuing to 28, the largest possible score. The scale runs as follows:

T-SCALE, WILSON LANGUAGE ERROR TESTS

Number correct answers	T-scale value	Number correct answers	T-scale value
0			
1		15	37
2	19	16	38.5
3	22.5	17	40.5
4	25	18	42.5
5	26	19	44.5
6	27.5	20	46
7	28.5	21	48
8	29.5	22	50
9	30.5	23	52.5
10	31.5	24	55.5
11	32.5	25	59
12	33	26	63
13	34	27	68
14	35.5	28	74

The T-scale gives a better comparative result for use with children and teachers, as it shows better the relative value of different scores.

The following table shows the T-scale score corrections for different ages and the resulting B-scores:

T-SCALE SCORE CORRECTIONS AND RESULTING B-SCALE (MEDIANS)

	Age 7	Age 8	Age 9	Age 10	Age 11	Age 12	Age 13	Age 14	Age 15	Age 16
1. T-scale scores (median).....	35.5	38.5	40.5	44.5	48	50	50	52.5	55.5	59
2. Correction.....	36	18.5	11.5	6.5	1.5	0	−.5	−4.5	−7.0	−11.5
3. B-score (median).....	71.5	57	52	51	49.5	50	49.5	48	48.5	47.5

Teachers unfamiliar with the T-scale and B-scale technique should continue the use of the raw scores. As teachers are becoming familiar with the language error tests, they are appreciating that their chief value is diagnostic and remedial, and for this purpose any method of scoring is satisfactory, as the main point is to make the pupils conscious of the particular errors which they need to correct.

NOTES ON ARTICLES IN EDUCATIONAL PSYCHOLOGY IN CURRENT ISSUES OF OTHER MAGAZINES

REPORTED BY CECILE COLLOTON

Department of Educational Psychology, The Lincoln School of Teachers College

Intelligence Tests

Intelligence Tests and Democracy in Education. Thomas M. Thompson. Educational Review, 1924, January, 5–11. Discusses the purposes and uses of tests and their importance in increasing social efficiency and real democracy.

The Rhode Island Intelligence Test. Grace E. Bird. Journal of Educational Research, 1923, December, 397–403. Describes in detail a test designed for children between the ages of three and six. Can be used either as an individual or group test. In process of standardization.

Intelligence Tests at Work. Elizabeth Kenney. The Journal of Educational Method, 1923, December, 147–152. Describes the use of the Otis test and teachers' judgments in classifying junior high school pupils.

Intelligence Tests and the Classification of Students in Ninth-grade English. Paul U. Sangren. Educational Administration and Supervision, 1923, December, 547–553. Terman group test of mental ability and achievement tests in English compared as bases of classification. Equally satisfactory results.

A Comparison of the Detroit First Grade Tests Given in Italian and English. Giovanni Giardini and W. T. Root. The Psychological Clinic, 1923, May–June, 101–108. Data from 340 Italian children do not show language difficulty to be an important factor in the rating of intelligence of foreign children. General cultural adaptation possibly of more influence.

The Witmer Formboard First Trial Records. Herman H. Young and Mary H. Young. The Psychological Clinic, 1923, May–June, 85–91. States the method of giving the Witmer Formboard test, discusses the index of formboard ability, and reports the first trial time records of 3622 cases.

The Diagnostic Value of the Audito-vocal Digit Memory Span. Anna S. Starr. The Psychological Clinic, 1923, May–June, 61–84. Describes the methods used and the results obtained in an investigation of 2000 children, ages 4 to 15. Norms characteristic of age and competency levels are given.

A Note on the Digit Test. Irene Case Sherman. The Psychological Clinic, 1923, May–June, 124. Failure of a subject nine years or older, chronologically, to repeat five digits is a good index of feeblemindedness.

Educational Tests

Difficulty, Reliability, and Grade Achievements in a Test of English Vocabulary. E. L. Thorndike and Percival M. Symonds. Teachers College Record, 1923, November, 438–445. Discussion of the practice effects and the reliability of four

alternative forms of an English vocabulary test used in the Classical League investigations.

An Educational Test for the Ballot. John R. Voorhis. Educational Review, 1924, January, 1–4. Discusses illiteracy in the United States and describes the New York State Regents Literacy Test administered by public school teachers to all new voters in the State.

The Principle of the Single Variable in a Speed of Reading Cross-out Test. J. Crosby Chapman and S. Cook. Journal of Educational Research, 1923, December, 389–396. A criticism of the Ayres-Burgess Silent Reading Scale and a description of the Chapman-Cook Speed of Reading Test, which, it is claimed, has all the merits which the Burgess scale lacks.

Tests for English Teachers. Allan Abott. The English Journal, 1923, December, 663–671. Describes a test for minimum essentials in the preparation of an English teacher and reports results from four steps of the preliminary test.

Putting Test Scores in a Form to Facilitate Comparisons. T. V. Goodrich, Journal of Educational Research, 1923, December, 377–388. Describe a method of reducing scores in original scales to terms of a scale based on grade norms in which the unit is a year's progress.

Miscellaneous

Criteria of Educational Research. Fowler D. Brooks. School and Society, 1923, December 22, 724–729. Part I gives seven standards for evaluating educational research. Part II discusses questionable procedures and practices in (*a*) planning and prosecuting research, (*b*) in collecting and evaluating data, and (*c*) in interpreting data and forming conclusions.

The Psychological Analysis of Spelling. Edith L. Hilderbrant. The Pedagogical Seminary, 1923, December, 371–381. Points to be considered in planning a method for improving spelling. Discussion based on the testimony of 600 high school students who cooperated to increase their own spelling efficiency.

Obstacles to Good Handwriting. J. B. Shouse. The Elementary School Journal, 1923, December, 301–306. Analyzes and discusses nine major mental difficulties in the way of improving handwriting.

A Study of Initial Stages in Reading by Pre-school Children. Arthur I. Gates and Eloise Bocker. Teachers College Record, 1923, November, 469–488. Reports the method used and the results obtained in a study of the learning of words by a group of children without previous experience. Eight tables give complete statistical data.

The Acquisition of Motor Control in Writing by Pre-school Children. Arthur I. Gates and Grace Taylor. Teachers College Record, 1923, November, 459–468. Reports an experiment in which two equivalent groups of pre-school children learn to write, one by tracing—one by copying letters with the advantage decidedly in favor of the copying group.

Teaching the Class and Reaching the Pupil. Frederick S. Breed. School and Society, 1923, December 15, 691–696. The need for educational reorganization to provide for individual differences and varying capacities.

The Superior Child. Alice M. Jones. The Psychological Clinic, 1923, May–June, 116–123. Four case studies of children of unusually high IQ who display talent along literary lines.

The Mental Tests of a Superior Child. H. Dealey Dvorak. Mental Hygiene, 1923, April, 250–257. Relates the story of a school official who refused to accept objective evidence which conflicted with his *a priori* judgment.

Mental Changes after Removing Tonsils and Adenoids. Gladys M. Lowe. The Psychological Clinic, 1923, May–June, 92–100. Tests of mental alertness, teachers' estimates of certain traits and scholarship records used with 35 children before and after operation for tonsils and adenoids show little or no improvement in general, over results from the control group. Confirms evidence from four other investigations.

A Comparison of a Group of High School Failures with a Group of Successful Students. T. V. Goodrich and S. L. Clements. School and Society, 1923, December 15, 715–720. Reports a study of 99 boys who were doing unsatisfactory work in high school and 74 boys who had made a semester grade of 80 per cent or above. Comparisons were made on basis of scores from standard tests and answers to questions relating to reasons for failure, amount of home study social life, vocational interests, etc.

A Graphic Rating Scale for Teachers. Max Freyd. Journal of Educational Research, 1923, December, 433–439. Describes a scale for rating teachers designed to cover, not the technique of teaching, but the psychological traits probably underlying success in teaching.

Proposals for Standardizing Measurement in Education. (1) *The Variability Unit.* Harry S. Will. Journal of Educational Research, 1923, December, 414–420. The Kental method as a unit for commensurating mental and educational measurements and facilitating comparison of teachers' marks with the results of standard tests. This method was first described in the February 1922 issue of the Journal of Educational Research.

The Otis Correlation Chart. Arthur S. Otis. Journal of Educational Research, 1923, December, 440–448. A detailed explanation of the calculation of the coefficient of correlation by means of this chart and the advantages in great ease and accuracy of computation.

How the High School Student Responds to Different Incentives to Work. Donald A. Laird. The Pedagogical Seminary, 1923, December, 358–365. Ninety-two college students report from memory their responses to sixteen different incentives used by their high school teachers. The relative merits of the various incentives are discussed in detail and all data are summarized in one table.

How the College Student Responds to Different Incentives to Work. Donald A. Laird. The Pedagogical Seminary, 1923, December, 366–370. College students report from memory their responses to 15 incentives which are discussed in detail. One table gives summarized data.

The Effect of Outside Work and Athletics upon Scholarship. D. A. Worcester. School and Society, 1923, December 29, 779–780. A study made at Kansas State Teachers College indicates that a student can do outside work and not endanger his scholarship standing but that the addition of a third activity, athletics, seriously affects his grades.

A Study of Adolescent Friendships. Paul E. Williams. The Pedagogical Seminary, 1923, December, 342–346. The answers of 84 delinquent boys, ages 12 to 17 years, show for those boys, the influence of age in the choice of friends, the qualities desired in a friend, ideas of loyalty toward friends, and the effect of cowardice on friendship.

NEW PUBLICATIONS IN EDUCATIONAL PSYCHOLOGY AND RELATED FIELDS OF EDUCATION

DEPARTMENT IN CHARGE OF LAURA ZIRBES[1]

An Objective Study of Sexual Interests

Adolescent Interests, A Study of Sexual Interests and Knowledge of Young Women, by F. I. Davenport, Archives of Psychology, No. 66. New York, 1923. Pp. 62. (Obtainable by addressing Substation 84, New York City.)

Utter disgust with subjective methods and a conviction that introspection might well be superseded by properly arranged and adapted objective psychological methods, have born fruit in a monograph which not only furnishes a wealth of data, but also demonstrates the effectiveness and straightforwardness of a new technique in the investigation of sex interests.

The investigator collected 880 spontaneous written questions submitted by a group of 160 young women in a city training school, prior to a course in a sex instruction for which the students had made application. In spite of the fact that all were high school graduates who had studied biology for at least one year, the study reveals a degree of ignorance so profound as to make intelligent interest in problems of sex hygiene impossible.

The interests or curiosities were classified after a careful analysis of the questions into "one term," "two term," and "three term" questions. The thirteen subdivisions are grouped under the following heads: (I) Physiology; (II) Reproduction; congenital conditions; childbearing—"racial" interests; (III) Sexuality—primitive sexual interest; (IV) Acquired Sexual Interests. About 78 per cent of the whole number of questions were concerned with information which properly belongs in the fields of hygiene and biology but in the thoughts of the questioners these matters are sexual. More than 6 per cent of the questions had their origin in sex superstitions. There was far less interest manifest in appropriate attitudes toward sexual life than in the reasons for shame or fear relative to aspects of sex. There was more evidence of interest in sex superstitions than in marriage.

[1] All unsigned reviews were prepared by Laura Zirbes.

The investigator concludes that the emphasis on primitive sexuality in society is reflected in the preponderance of questions of that sort, and that the failure of society to provide positive, adequate, and specific training for acquired and socially approved forms of sex interest tends to accentuate the primitive at the expense of higher sexual sentiments and interests into which the former may be developed (or sublimated) through education.

There are those who stoutly maintain and defend taboo as a social and psychological expedient. They should contemplate the probable misinformation which would be dispensed by these young women should they later proceed to instruct their charges in school and home, with no further education than that provided by first hand experience and incidental contacts. Surely society needs the intelligent cooperation of its youth in the problems of sex conduct.

Because the most serious problems are problems of conduct, the dispensing of specific sex information is neither the crux nor the cure-all of the situation. This study should be extended. The author suggests that similar data be gathered from young men for a comparative study. The social implications of such findings should be an impetus and guide to a positive educational program, conceived with the conviction that society is concerned and responsible, because habits and attitudes are so highly modifiable by environment, experience and education.

An Inventory of the Minds of Young Children

An Inventory of the Minds of Individuals of Six and Seven Years Mental Age, by Grace A. Taylor. Teachers College Contributions to Education, No. 134. New York, 1923. Pp. 144.

An analysis of the knowledge and skills of six and seven-year-old children is of equal interest to the first grade teacher and to the teacher of feeble-minded children. The purpose of the study, as defined by the author, is the determination of certain things which are commonly known by children at the time of school entrance, and a comparison of the responses of normal and bright children of six and seven year mentality with feeble-minded children of the same mental age. Three groups of children, selected upon a basis of individual mental test rating, are considered; institutional feeble-minded, public school children of normal age for the first grade, and superior kindergarten children. The materials used in the inventory include individual

and group performance tests of wide variety, and questions and exercises pertaining to personal and general information, knowledge of social and moral situations, memory, number, and numerous other items of knowledge or skill.

The findings are important from the standpoint of educational procedure both in the school and the home, and are of especial significance to the student of mental tests in that many of the common test elements are objectively analyzed. Upon the whole, the findings would suggest that there is decided agreement in the items which are generally known or unknown to children of six or seven year mentality, regardless of chronological age, and that such differences as may exist seem to be due to biological growth, years of life experience, and innate ability to organize experiences.

BESS V. CUNNINGHAM.

Teachers College, N.Y.C.

A RESPONSE TO THE DEMAND FOR NEW LIGHT ON THE TEACHINGS OF PHONICS

How to Teach Phonics, by Mary L. Dougherty. New York: Houghton Mifflin Company, 1923. Pp. XI–89.

This book will no doubt be welcomed by those who are impatient for authoritative advice on practical problems of instruction in reading. It is well sponsored. The author, herself an instructor in education at Johns Hopkins University, acknowledges her indebedtness to Professor Frank N. Freeman for careful reading and helpful criticism of the manuscript. Dr. Henry Suzzalo contributes a six-page editorial introduction in which the history and status of instruction in phonics are summarized, and the place and purpose of such instruction are clearly set forth. In the first chapter the author says:

"This book has been prepared with the common reading vocabulary of many books in mind, so that it may easily be used with any reading book suitable for beginners and with later books satisfactory for the primary grades . . . It is the purpose of this book to suggest a middle course, giving a minimum amount of training consistent with efficiency . . . As an end in itself, phonics has no place in the schoolroom . . . and no amount of such skill *alone* will result in efficient reading."

The same chapter emphasizes the close relationship between the use of phonics and thought getting, and stresses the dangers of over- or underemphasis on mechanics in beginning reading.

From this point on, the tone of finality and the dogmatic assumption of authority should suffice to quash any argument or opposition. Most teachers will be glad to be told just what to do and how to do it. But the reviewer, who happens to be an investigator in reading must assume the role of "higher critic." Anything so didactic and prescriptive must reveal the sources upon which it proceeds, and these sources must be scientifically sound, objective and sufficient to lead others to similar conclusions and similar practical interpretations. The principles or assumptions upon which prescribed methods rest should at least be stated (1) in order that, when specific "devices" fail, the principle may be reinterpreted in a revised procedure: (2) In order that the principle itself may be critically examined, or held as a tentative hypothesis requiring proof: (3) In order that unwarranted assumptions may not lead practically inclined persons to an uncritical revision of practice: (4) In order that the relative significance and soundness of each suggestion may be made to depend on its own supporting data: (5) In order that alert and effective teachers may have criteria against which to check their own original procedures, and be free to work at the perfection of their art: (6) In order that young andin experienced teachers have that opportunity for professional growth which inheres in the conscious search for principles, the rationalization of procedures and the contemplation of objective evidence from which truly authoritative advice proceeds. Because of this evident deficiency the reviewer for one can only use the book as problem material requiring such critical validation.

Some of the "devices" are ingenious; some are inexcusably inartistic and unnecessarily formal.

There is nothing to indicate that the author has considered it necessary to go beyond her own ingenuity in the construction of methods and nothing to show that the proposed methods have stood the test of use or proved themselves effective in the tested results of classroom experimentation. School systems which adopt the course as prescribed would do a real service if they could arrange experiments which might better have preceded publication.

The book contains no bibliography. The total number of references cited in the body of the text and in footnotes follow in the following inventory: Two page references and two brief quotations from Dr.

Judd's monograph entitled "Reading: Its Nature and Development." One brief quotation and reference to an article by Dr. W. S. Gray, Elementary School Journal, Volume XIX pp. 511–512, one page reference and one brief quotation from Dr. Freeman's Psychology of the Common Branches, one quotation and reference to Professor S. C. Parker's "Types of Elementary Teachers and Learning," two references to Packer's vocabulary study in the Twentieth Yearbook and two references to the "The Jones Scale of Reading." There is no other evidence of an intimate acquaintance with the extensive scientific literature of the field.

Psychology for Teachers

Everyday Psychology for Teachers, by F. E. Bolton. New York: Charles Scribner's Sons, 1923. Pp. 443.

Of the four parts into which Professor Bolton has divided his book, Part III alone, dealing with Modes of Learning and Behavior, is likely to attract more than superficial attention. The three other parts, dealing with the importance of the science of education; the meaning and end of education; the rationale of effective study; native endowment—differences, heredity, instinct; and measurement, are too compressed and too sketchy to satisfy anything but a professorial conscience for leaving nothing unmentioned.

Part III, which includes more than half of the four hundred odd pages of the book, presents up-to-date and stimulting discussions of the various aspects of broadly conceived and socially directed learning. Sense-perceptions, play, the emotions, interest, memory, imitation, thinking, the will, and formal discipline are respectively defined, illustrated, their educational significance and limitation noted, and their recognition in method at least partially suggested. Particularly well done are the treatments of Learning and Expression through Play, The Education of the Emotions, Motivation and Initiative, Memory, Learning through Imitation, Imagination in Learning and Expression, and The Direction of Behavior.

The points of view displayed are easily traceable to the three men whose influence the author specifically acknowledges in his preface—G. Stanley Hall, William James, and John Dewey. The social imperative of an educational method determined by reference to the nature of the child, the wide pedagogical and moral significance of motor activity and play, the fundamental interrelations of interest

and effort, the necessity of identifying and considering stages of individual development, the universal application of the habit concept, and the deep ramification of the affective phase of experience—these ideas in turn have the right of way through Professor Bolton's book. They are, of course, very familiar to even the moderately advanced student of education, but it is not to him or his superior that this book is really directed.

"This book has been written," says the author, "for... the young teachers who wish to get a right start in their profession by becoming acquainted with the principles of the subject; the old teachers who wish to have the result of experiment and research that have met with most favor; and that great number of persons interested in the subject of education who have faith in the boys and girls of to-day and wish to have a part in making them the worthy citizens of to-morrow." This audience, though not so homogeneous possibly as to be all satisfied with the same book, will doubtless be pretty well pleased in going through this one to note the absence, or relative infrequency, of crowded data tables, statistical formulas, graphical but, nevertheless, obscure illustrations, detailed accounts of experimentation, and any scientifically approved but, to the general reader, pusillanimous reticence in stating conclusions or prescribing authoritatively what is to be done about things. The book is clear enough to be understood, and definite enough to be helpful.

In perhaps only one important particular will the readers for whom the book is intended be disturbed. They will probably object to the author's rather generous strewing about of quotations in certain chapters. The effect of this excessive summoning of sums at times is not to intensify the author's own radiations, but to eclipse them and to reduce him to little more than an irresolute oscillation on the page. Meanwhile with so much light on the subject, from so many rapidly succeeding sources, the reader grows weary of adjusting his eyes and his mind and begins to hunt for illumination more steady and more filtered. The author escapes this quoting habit, however, in the body of the book and says most of what is said about behavior and learning in his own really effective way.

The book is suitable for use in training school classes. The chapters are provided with sets of "Suggestive Questions" and brief bibliographies of general works.

M. H. Willing.

The Lincoln School of Teachers College.

MAR 20 1924